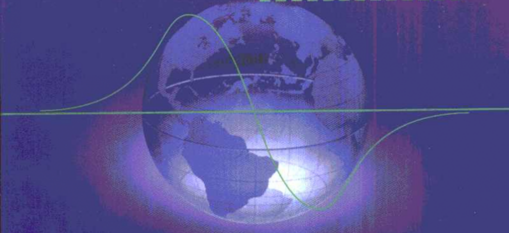
中国地质大学"创新工程"丛书

地球物理反演理论

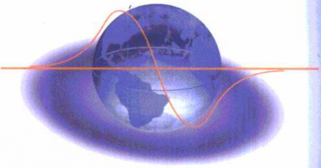
王家映 编著



中国地质大学出版社

CUGP

责任编辑：刘先洲
封面设计：田剑云



ISBN 7-5625-1375-9

9 787562 513759 >

定价：10.00元

地球物理反演理论

王家映 编著

中国地质大学出版社

内容提要

本书系统地介绍了在地球物理学中广泛应用的线性反演理论及正在兴起的非线性反演方法。全书共五章，第一章：线性反演理论概述；第二章：线性问题的长度解；第三章：广义反演法；第四章：Backus-Gilbert 反演理论；第五章：非线性反演方法。前四章属线性反演理论，第五章简单介绍了目前常用的一些非线性反演技术。

本书既有各种反演方法的数学公式推导，又有物理概念分析，易于理解和掌握，具有较强的理论性和实用性，可供地球物理探测与信息技术专业的研究生和本科生以及其他从事地球物理、地球化学、数学地质、遥感等工作的科技人员和研究人员参考。

图书在版编目(CIP)数据

地球物理反演理论/王家映编著．—武汉：中国地质大学出版社，1998.12

ISBN 7-5625-1375-9

Ⅰ.地…

Ⅱ.王…

Ⅲ.地球物理学-B-G 理论-反演计算方法

Ⅳ.P3

出版发行　中国地质大学出版社（武汉市喻家山．邮政编码 430074）
责任编辑　刘先洲　　责任校对　胡义珍　　版面设计　阮一飞
印　　刷　武汉测绘院印刷厂
经　　销　湖北省新华书店

开本 850×1168　1/32　印张 6　字数 160 千字
1998 年 12 月第 1 版　1998 年 12 月第 1 次印刷　印数 1—1 000 册
定　价：10.00 元
ISBN 7-5625-1375-9/P.497

前　言

宇宙万物都在矛盾中存在着、运动着。任何事物都是一分为二的，对立统一的，既有正面，也有它的反面。在初等数学中有指数和对数，三角函数和反三角函数，双曲函数和反双曲函数等；在高等数学中有傅氏变换和傅氏反变换，拉普拉斯变换和拉普拉斯反变换，褶积和反褶积等；在物理学中有正电子和负电子，质子和反质子，物质和反物质等；同样，在地球物理学中有正(演)问题和反(演)问题。

如果我们用已知自变量 x 去求函数 $y=f(x)$ 的值，已知一个函数 $f(x)$ 去求其变换 $F=L(f(x))$，已知一个地球物理模型去求其响应函数这类问题叫正演问题的话，那么由 $f(x)$ 求 x，由 $L(f(x))$ 求 $f(x)$，由地球物理观测资料去反推地球物理模型，这类问题就统称为反问题，或者求反演问题。

反问题是对正问题而言的。比如，微分方程的正问题是研究如何描述与刻画物理过程、系统状态等现象(即建立微分方程)，以及根据过程与状态的特定条件(初始或边界条件)去求解这一定解问题，从而得到过程与状态的数学描述；而微分方程的反问题，是指从微分方程解的某些泛函去确定微分方程的系数(或其右端项)。很明显，这里有两类反演问题，第一类是确定过程的过去状态；第二类是借助解的某些泛函去确定微分方程的系数。

无论哪一类反问题，都是属于地球物理反演理论研究的范围。因此，“剔除”具体反演问题的物理内容，地球物理反演就可以概括为：地球物理学中的反演理论就是研究把地球物理学中的观测数据映射到相应的地球物理模型的理论和方法。

地球物理学涉及广泛的内容，有地震学、重力学、电磁学、热学、

地球年龄学等，即使在同一学科分支中也有各种不同的资料采集方法，如地震学中就有反射法和折射法、电磁法中有直流电法和交流电法(或电磁法)等。在各种方法中，又有许多变种，如交流电法中有人工源和天然源电磁法之分等等。尽管地球物理学家研究地球所依据的物性参数不同，方法各异，正演公式也千差万别，但是观测资料的反演方法却有许多共同之处。作为反演理论，本书只涉及各种地球物理观测数据反演方法之共同理论，反演中所遇到的共同问题，以及解决这些问题所必须采取的共同措施(假定读者已经熟悉各种地球物理的正演问题)。然而，为说明各种反演方法的实际效果，我们又不得不以某些具体地球物理问题为例，从而也不可避免地会涉及一些具体的正演问题。

绝大多数地球物理问题都是非线性问题。然而，在一定条件下非线性问题可以线性化，即把非线性问题化成线性问题。众所周知，地球物理反演问题有线性反演和非线性反演之分，前者指观测数据和地球物理模型之间存在线性关系(线性函数或线性泛函)，而后者是非线性关系(非线性函数或非线性泛函)。线性反演法已经形成了完整、系统的理论，近 30 年来获得了广泛的应用。非线性反演法近年来发展迅速，日新月异，已成为反演理论家重点研究的领域。按照在实现反演映射时是否需要迭代，可把反演方法分为间接反演法和直接反演法，需要迭代的叫间接反演法，反之称直接反演法；按照应用的领域，人们又常把反演方法分为诸如地震反演法、重力反演法等，甚至有人以反演方法的提出人命名，如高斯-牛顿法，马夸特法等。

虽然反演理论已在地球物理的各个领域获得了广泛而成功的应用，但它毕竟是近 30 年才发展起来的地球物理学的新兴分支，还不能说它的理论(特别是非线性反演理论)已经完备，效果令人十分满意。和任何新兴学科或新生事物一样，它还有许多问题亟待解决，就是现有的方法和理论也还需要在实践中不断改进、补充和完善。

应该指出，既不应该把地球物理学中的反演理论当成与地球物

理问题无关的纯抽象的数学问题来学，也不应该把它看成是一种纯方法技术问题，忽视其理论性和一般指导原则，否则就是片面的。笔者认为，只有具备一定的数理基础，同时又较好地懂得地球物理学的人，才能更好地完成反演工作所面临的任务。

如果我们把地球物理问题分为资料采集、数据处理和反演解释三个阶段的话，那么，资料采集是基础，数据处理是手段，反演解释才是地球物理工作的目的。反演解释工作是地球物理工作中最重要的一个阶段。随着地球物理工作的不断深入和发展，随着科学技术特别是计算技术的发展，反演理论会日臻完善，会越来越显示出强大的生命力，为当今地球科学的发展作出自己的贡献！

第一章　线性反演理论概述

§1　反演理论的目的和任务

目前，人类对地球内部的物理性质（包括速度、密度、电导率、温度等）以及矿产资源的分布已经有了不同程度的了解。这种知识多数来自于地表地质和地球物理、地球化学资料的反演和解释，而不是来自于钻井。对于地球表层的了解是如此，深层更是如此。通过大陆深钻和海洋深钻获取地壳深处的地质、地球物理信息是必要的，但是它毕竟耗资巨大，而且深度有限，所获得的资料也仅来自一孔之见。要少花钱多办事的唯一途径是应用地球物理和地球化学的方法，在采集到可靠的第一手资料后，加强资料的处理和反演解释，以获取更接近于真实的地球物理和地球化学模型！

由此可见，地球物理学中的反演理论的目的是根据观测数据求取相应的地球物理模型。所以，首先必须确定观测数据和地球模型参数之间的函数关系，使地球物理工作者既可根据给定的模型参数计算相应的观测数据（即实现正演计算），也可以根据观测数据求取地球物理模型的参数，实现反演映射。显然，正演是反演的前提和条件，只有解决了正演计算，不管是靠解析的方法还是数值的方法，才有可能实现反演映射。遗憾的是，并不是对所有地球物理问题，科学家都弄清了它们的机理，确定了它们的数学物理模型的，如天然地震预报、地磁场的起因和向西漂移等就是这类问题的典型代表。毫无疑问，对这类问题，目前反演是无能为力的。

但是，这并不是建立了正确的数学物理模型以后，反演问题就完全解决了。那种把反演问题和正演问题等同起来，以为反演理论没有

自身需要解决的特殊问题，显然也是不对的。

和其他学科一样，反演理论也有其需要解决的特殊问题。正如著名的反演理论家 R. Parker 在其有名的论文《Understanding Inverse Theory》中所概括的那样，反演理论必须解决四大问题。

(1)解的存在性：即给定一组观测数据后，是否一定存在一个能拟合观测数据的解或模型；

(2)模型构制：如果存在性是肯定的，如何求得或构制能拟合观测数据的模型；

(3)非唯一性：能拟合观测数据的模型是唯一的，还是非唯一的；

(4)结果的评价：如果解是非唯一的，如何才能从构制的模型中提取关于真实模型的地球物理信息。

在地球物理资料的反演中，解的存在性已被大量的理论文章和实际资料所证实。地球物理学家把存在性问题留给数学家们去研究，他们对此并不十分感兴趣。然而，这并不是说这个问题毫无意义。事实上，它是对所研究的地球物理问题的数学物理模型及其假设条件的正确性的检验，具有重大的理论意义和明显的实际价值。

关于模型构制，解的非唯一性和结果的评价，是本书要讲解的主要内容，以后几章中要详细研究。

值得提及的是，有的反演理论家将解的稳定性问题也作为反演理论必须研究的重要问题之一，我们也将在适当的章节中加以介绍。

§2 数学物理模型和响应函数的正演问题

在地球物理学中，将观测数据和地球的物理模型参数联系起来的数学表达式叫数学物理模型。不同的地球物理问题，其数学物理模型是不同的，就是同一个地球物理问题，其观测方式不同，近似条件有变化，其地球物理模型也不一样。

虽然地球物理问题千差万别，但把观测数据和物理模型参数联系起来的数学表达式却只有线性和非线性两大类。如以 x 表示模型

参数，y 表示观测数据，F 表示联系 x 和 y 的函数或泛函表达式，则满足

$$
\begin{aligned}
&(1) \quad F(x_1+x_2)=F(x_1)+F(x_2)=y \\
&(2) \quad F(\alpha x)=\alpha F(x)=y
\end{aligned}
\tag{1.1}
$$

两个条件时，称 F 为线性函数或线性泛函，其中 α 为常数。

不言而喻，凡是不满足(1.1)式的函数或泛函就是非线性的了。在地球物理学中，绝大多数观测数据和模型参数之间都不满足线性关系。但是，在一定近似条件下均可简化或近似简化为线性关系。因此，线性反演问题是地球物理学家最关心的问题之一，也是本书要论述的重点。

不管是在线性反演法中，还是在非线性反演法中，都涉及到地球响应函数(或理论观测值)的计算。正演是反演的前提和条件，只有准确地计算出地球的响应函数，才有可能求得可靠的地球物理模型。

如果观测数据和地球物理模型之间存在着确定的函数关系，这种正演计算在计算机程序中是不难实现的。但是，在许多情况下，如二维、三维时，观测数据和地球物理模型之间并不存在明显的函数关系，这时就应借助于数值的方法，比如用有限差分，有限单元，积分方程等方法来实现正演计算。

§3 非线性问题的线性化与连续模型的离散化

如前所述，将非线性问题线性化，使非线性反演问题简化为线性反演问题，是当今解决非线性反演问题的一个重要途径和方法。下面将简单介绍一些非线性问题线性化的方法。

1. 参数置换法

所谓参数置换法，就是通过参数置换将非线性的地球物理方程线性化，以便于应用线性反演方法求解模型的参数。

例 1 某地[其坐标为(x,y,h)]发生天然地震后，在其附近的地震观测站[其坐标为$(x_k,y_k,0)$]观测到直达波 P 的走时 t_{Pk}，这里 $k=$

$1,2,\cdots,n$；n 为观测站的数目。假定地球介质是均匀的，发震时刻为 t_0，则下列方程存在：

$$(x-x_k)^2+(y-y_k)^2+h^2=(t_{Pk}-t_0)^2V_P^2 \qquad (k=1,2,\cdots,n)$$

这里 V_P（或 v_P）为介质中地震波的传播速度。

对地震参数(x,y,h,t_0,V_P)而言，上式是一个非线性方程。

设：$a_{k1}=1$，

$a_{k2}=-2x_k$，

$a_{k3}=-2y_k$，

$a_{k4}=-t_{Pk}^2$，

$a_{k5}=2t_{Pk}$，

$b_k=x_k^2+y_k^2$

以及

$\rho_1=x^2+y^2+h^2-t_0^2V_P^2$，

$\rho_2=x$，

$\rho_3=y$，

$\rho_4=V_P^2$，

$\rho_5=t_0V_P^2$。

则上式可化为

$$a_{k1}\rho_1+a_{k2}\rho_2+a_{k3}\rho_3+a_{k4}\rho_4+a_{k5}\rho_5+b_k=0 \qquad (k=1,2,\cdots,n) \tag{1.2}$$

由于以上方程的系数矩阵$\{a_{ki}\}$和常数项 b_k 可以根据第 k 个观测点的地理坐标(x_k,y_k)和地震波的走时 t_{Pk}求得。因而，只要 $n>5$，则 $\rho_1,\rho_2,\rho_3,\rho_4,\rho_5$ 就不难用线性反演理论求得，从而地震参数(x,y,h,t_0,V_P)就可迎刃而解了。

例 2 用统计法求取弹性介质的吸收系数。

大家知道，从震源发出的地震波在到达接收点时，其振幅会发生变化。在无色散系统中，振幅的变化主要来源于波前面的扩散和介质

的吸收。前者与距离 r 成反比，后者与距离 r 呈负指数的关系，即

$$A_i = cr_i^{-1}e^{-\alpha r_i}$$
$$= at_i^{-1}e^{-bt_i} \qquad (i=1,2,\cdots,n)$$

式中：c 是比例常数；$a=cv^{-1}$，$b=\alpha v^{-1}$，α 是吸收系数。

从上式不难看出，观测数据 A_i 和待求系数 a 和 b 之间是非线性关系。但是，如将上式两端同取对数，则有：

$$\ln A_i = \ln a - \ln t_i - bt_i \qquad (i=1,2,\cdots,n)$$

或

$$y_i = u - bt_i \qquad (i=1,2,\cdots,n)$$

其中：

$$y_i = \ln A_i + \ln t_i \qquad (i=1,2,\cdots,n) \tag{1.3}$$
$$u = \ln a$$

只要 $n>2$，解线性方程组(1.3)，不难求得 u 和 b，因而就很容易求出 c 和吸收系数 α 了。

2. 台劳级数展开法

级数理论告诉我们，任何一个函数 $f(x)$，如果满足条件：

(1)在点 a 的某邻域 $|x-a|<\delta$ 内有定义；

(2)在此邻域内从 1 阶一直到 $(n-1)$ 阶的导数 $f'(x)$，…，$f^{n-1}(x)$存在；

(3)在 a 处有 n 阶导数 $f^{(n)}(a)$，那么 $f(x)$在点 a 的邻域内可表成如下台劳级数：

$$f(x)=f(a)+f'(a)(x-a)+\frac{1}{2!}f''(a)(x-a)^2+\cdots+\frac{1}{n!}f^{(n)}(a)(x-a)^n+o(|x-a|^n)$$

其中：$o(|x-a|^n)$为高阶无穷小。如果仅取前两项作为 $f(x)$的一阶近似，则有：

$$f(x)=f(a)+f'(a)(x-a)+o'(|x-a|) \tag{1.4}$$

忽略上式中的一阶无穷小，则有：

$$f(x)\approx f(a)+f'(a)(x-a) \tag{1.5}$$

如果 $\boldsymbol{x}$ 是 N 维向量，则有：

$$f(\boldsymbol{x})=f(\boldsymbol{a})+\sum_{i=1}^{N}\frac{\partial f(\boldsymbol{a})}{\partial a_i}(x_i-a_i) \tag{1.6}$$

式中：

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_N\end{bmatrix};\qquad \boldsymbol{a}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_N\end{bmatrix}$$

将(1.6)式应用于非线性地球物理问题，则有：

$$d_j=d^\circ_j+\sum_{i=1}^{N}\left(\frac{\partial f_j}{\partial m_i}\right)^\circ\Delta m_i,\qquad (j=1,2,\cdots,M)$$

或

$$\Delta d_j=\sum_{i=1}^{N}\left(\frac{\partial f_j}{\partial m_i}\right)^\circ\Delta m_i \tag{1.7}$$

式中：d_j 为第 j 个观测数据，且

$$d_j=f(\boldsymbol{m},\lambda_j)\qquad (j=1,2,\cdots,M)$$

是一个非线性函数，$\boldsymbol{m}$ 是 N 维模型向量；$\left(\frac{\partial f_j}{\partial m_i}\right)^\circ$代表在起始模型 $\boldsymbol{m}^\circ$ 处，第 j 个观测值 f_j(或 d_j)对第 i 个模型参数 m_i 之偏导数；$\Delta m_i=m_i-m^\circ_i$，即第 j 个模型参数之增量；$\Delta d_j=d_j-d^\circ_j$ 之差，是第 j 个观测数据 d_j 和在 $\boldsymbol{m}^\circ$处的理论数据 d°_j 之差，即观测数据之增量。

将(1.7)式改写为矩阵方程，即有：

$$\Delta\boldsymbol{d}=\boldsymbol{G}\Delta\boldsymbol{m} \tag{1.8}$$

式中：

$$\Delta\boldsymbol{d}=\begin{bmatrix}\Delta d_1\\\Delta d_2\\\vdots\\\Delta d_M\end{bmatrix},\qquad \Delta\boldsymbol{m}=\begin{bmatrix}\Delta m_1\\\Delta m_2\\\vdots\\\Delta m_N\end{bmatrix}$$

和

$$G=\begin{bmatrix} \left(\frac{\partial f_1}{\partial m_1}\right)^\circ & \left(\frac{\partial f_1}{\partial m_2}\right)^\circ & \cdots & \left(\frac{\partial f_1}{\partial m_N}\right)^\circ \\ \left(\frac{\partial f_2}{\partial m_1}\right)^\circ & \left(\frac{\partial f_2}{\partial m_2}\right)^\circ & \cdots & \left(\frac{\partial f_2}{\partial m_N}\right)^\circ \\ \vdots & \vdots & & \vdots \\ \left(\frac{\partial f_M}{\partial m_1}\right)^\circ & \left(\frac{\partial f_M}{\partial m_2}\right)^\circ & \cdots & \left(\frac{\partial f_M}{\partial m_N}\right)^\circ \end{bmatrix}$$

由此可见，将非线性地球物理响应函数线性化，可以得到形如(1.8)所示的矩阵方程。到此，不难用线性反演理论，经反复迭代求出拟合观测资料的一个模型。

将(1.6)式应用于目标函数(或方差函数)，将非线性的目标函数线性化，也可利用线性反演理论，实现观测资料的反演。

设目标函数

$$\Psi=\sum_{j=1}^{M}\left(\frac{d_j-d^\circ_j}{d_j}\right)^2 \qquad (j=1,2,\cdots,M) \tag{1.9}$$

将它在初始模型 $\boldsymbol{m}^\circ$附近按台劳级数展开，并略去三次以上的高阶项，则有：

$$\Psi=\Psi^\circ+\sum_{i=1}^{N}\left(\frac{\partial\Psi}{\partial m_i}\right)^\circ\Delta m_i+\frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left(\frac{\partial^2\Psi}{\partial m_i\partial m_l}\right)^\circ\Delta m_i\Delta m_l$$

和以前一样，角标“°”仍表示在初始模型处之值。若将上式写成矩阵，则得

$$\Psi=\Psi^\circ+\Delta\Psi^{\mathrm{T}}\Delta\boldsymbol{m}+\Delta\boldsymbol{m}^{\mathrm{T}}\mathbf{Q}\Delta\boldsymbol{m} \tag{1.10}$$

式中：

$$\Delta\Psi=\begin{bmatrix}\left(\frac{\partial\Psi}{\partial m_1}\right)^\circ \\ \left(\frac{\partial\Psi}{\partial m_2}\right)^\circ \\ \vdots \\ \left(\frac{\partial\Psi}{\partial m_N}\right)^\circ\end{bmatrix}; \qquad \Delta\boldsymbol{m}=\begin{bmatrix}\Delta m_1\\ \Delta m_L\\ \vdots\\ \Delta m_N\end{bmatrix} \tag{1.11}$$

$$Q=\begin{bmatrix} \left(\frac{\partial^2\Psi}{\partial m_1\partial m_1}\right)^\circ & \left(\frac{\partial^2\Psi}{\partial m_1\partial m_2}\right)^\circ & \cdots & \left(\frac{\partial^2\Psi}{\partial m_1\partial m_N}\right)^\circ \\ \left(\frac{\partial^2\Psi}{\partial m_2\partial m_1}\right)^\circ & \left(\frac{\partial^2\Psi}{\partial m_2\partial m_2}\right)^\circ & \cdots & \left(\frac{\partial^2\Psi}{\partial m_2\partial m_N}\right)^\circ \\ \vdots & \vdots & & \vdots \\ \left(\frac{\partial^2\Psi}{\partial m_N\partial m_1}\right)^\circ & \left(\frac{\partial^2\Psi}{\partial m_N\partial m_2}\right)^\circ & \cdots & \left(\frac{\partial^2\Psi}{\partial m_N\partial m_N}\right)^\circ \end{bmatrix}\cdot\frac{1}{2} \quad (1.12)$$

目标函数(1.10)极值存在的必要条件是

$$\frac{\partial\Psi}{\partial\Delta\boldsymbol{m}}=0,$$

故(1.10)变为

$$\Delta\Psi=\boldsymbol{Q}\Delta\boldsymbol{m} \quad (1.13)$$

显然，(1.13)和(1.8)完全相同。应该指出，对线性方程(1.8)和(1.13)求解可以得到 $\Delta\boldsymbol{m}$，即求得参数的校正向量。它们还不是待求的模型参数 $\boldsymbol{m}$，还必须对初始模型进行校正，校正后的模型也不一定就是待求的模型(不一定能拟合观测数据)。因此，还必须以校正后的模型作为初始模型，重复上述步骤，反复迭代，直至拟合观测数据为止。

除以上两种非线性函数线性化的方法外，在地球物理资料反演中，还有许多种其他方法，读者可参考有关文献。

在连续介质情况下，如果观测数据与连续模型之间呈非线性泛函关系，也可以在一定近似条件下，用各种不同的方法，将非线性泛函化为线性泛函，这已成了近年来解决非线性泛函反演问题的重要途径。

其实，在观测数据和模型之间呈线性泛函，即当：

$$d_j=\int_0^\infty g(z,\tau_j)m(z)\mathrm{d}z \qquad (j=1,2,\cdots,M) \quad (1.14)$$

时，也可以将连续模型 $m(z)$ 离散化，实现观测数据的反演。(1.14)中 d_j 是第 j 个观测数据；$m(z)$ 是地球物理模型，是 z 的连续函数，它可

是电阻率、速度、密度、温度以及其他任何待求的地球物理参数，$g(z,\tau_j)$是核函数，它既是z的函数，也是参数τ_j的函数。

如将线性积分方程(1.14)中的模型$m(z)$离散化，将它分为N个区间，并假设每个区间中的模型$m_i(z)$是一个常量，则(1.14)可写为

$$d_j=\sum_{i=1}^{N}G_{ji}m_i \qquad (j=1,2,\cdots,M) \tag{1.15}$$

式中：

$$G_{ji}=\int_{z_{i-1}}^{z_i}g(z,\tau_j)\mathrm{d}z \qquad (j=1,2,\cdots,M) \tag{1.16}$$

显然(1.15)可以表示为矩阵方程，则有：

$$\boldsymbol{d}=\boldsymbol{Gm} \tag{1.17}$$

式中：

$$\boldsymbol{d}=\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{bmatrix}; \qquad \boldsymbol{m}=\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_N \end{bmatrix} \tag{1.18}$$

且

$$\boldsymbol{G}=\begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1N} \\ G_{21} & G_{22} & \cdots & G_{2N} \\ \vdots & & & \\ G_{M1} & G_{M2} & \cdots & G_{MN} \end{bmatrix} \tag{1.19}$$

由矩阵代数不难求得(1.17)式的解。这里，需要着重说明的是，(1.17)和(1.14)式是线性反演理论中最为重要的两个方程。线性反演理论就是研究解这两个方程的理论、方法，以及由此而引申的一系列问题。

§4 模型构制

现在将讨论模型构制，即平常所讲的“反演方法”问题，这是本书

将要研究的重点。在这一节中，我们将简要地介绍一下模型构制的基本思路和基本方法，让读者有一粗略了解，为后面几章的学习打下基础。

让我们回到(1.17)，且设

$$\underset{(M\times 1)}{\boldsymbol{d}} = \underset{(M\times N)}{\boldsymbol{G}}\underset{(N\times 1)}{\boldsymbol{m}}$$

其中：$\boldsymbol{d}$ 是 $M\times 1$ 维向量；$\boldsymbol{G}$ 是 $M\times N$ 阶矩阵；$\boldsymbol{m}$ 是 $N\times 1$ 维向量。显然，这里有四种情况：

(1)当 $M=N=r$，且 $\boldsymbol{G}^{-1}$ 存在时，方程(1.17)有唯一解。这里 r 是矩阵 $\boldsymbol{G}$ 的秩，而 $\boldsymbol{G}^{-1}$ 表示矩阵的逆；

(2)当 $M>N=r$ 时，方程(1.17)是超定方程，无常规意义下的解，但有最小方差解；

(3)当 $N>M=r$ 时，方程(1.17)是欠定方程，也无常规意义下的解，但有 $|\boldsymbol{m}|=\min$ 意义下的最小长度解；

(4)当 $\min(M,N)>r$ 时，方程(1.17)的问题是混定问题。这时，只有同时在两种限制条件(即最小方差和模型最小长度条件)限制之下，方程才有解。

这里讲的方差或模型长度都是指在 L_2 范数意义下的欧基里德空间的长度。当然，也可以在其他测度(L_1，L_3，…，L_∞)为最小的意义下估算模型参数。测度定义不同，对模型参数和观测数据加权方式不同，就会派生出各种不同的反演方法。

以上讨论了地球物理模型参数为离散情况下(包括将连续模型离散化以后的情况)模型构制的基本思路和基本方法。下面，将介绍在连续介质条件下如何反演。

介质的物性是空间坐标的连续函数，有以下三种情况：

(1) $m=m(z)$，即介质的物性仅是一个空间坐标，比如说深度 z 的函数，即一维模型；

(2) $m=m(x,z)$，即介质的物性是两个空间坐标 x、z 的函数，即二维模型。这时，我们常称水平坐标方向 y 为二维模型的走向，x 为

倾向。

(3)$m=m(x,y,z)$，是三维介质，这时介质没有明显的走向。实际的地球物理模型都是三维的，只是在一定近似条件下方可视为二维，甚至一维的。为什么要把实际地球物理模型近似为二维和一维的呢？因为一维、二维的正演公式简单，计算方便。这种近似，在较小的空间范围内和一定精度意义下是允许的。

在连续介质情况下，欲根据有限个观测数据 $d_j(j=1,2,\cdots,M)$，求取无限个未知数的连续函数，无疑是不可能的，必须附加一些已知的限制条件或先验信息，附加的条件越多、越严格，求得的模型越接近于真实情况。这一点，我们还要在以后的章节中详细加以论述。

由上面的讨论可以看出，在模型构制时，如何处理观测数据和模型参数，有三种不同的观点和方法：

第一，把观测数据和模型参数(既可是离散模型，也可以是能用有限个参数表征的连续模型)都看成是随机变量，通过研究它们所遵循的概率分布对观测资料进行反演。

第二，把观测数据视为随机变量，把模型看成是由一些确定参数所决定的。反演的任务就是估算这些待定的参数以及它们可能具有的误差，这是本书介绍的各种反演方法所持的主要观点。

第三，视观测数据为随机变量，模型为连续函数，形成一套连续介质的反演理论和方法，这就是著名的 Buckus-Gilbert 反演法，这是本书讨论的又一重点。

§5 解的非唯一性

正如前面所指出的，解的非唯一性是地球物理资料反演中最重要的问题之一。对所有地球物理问题都存在，只是程度不同而已。有的非唯一性问题小，有的非唯一性问题大。反演中的非唯一性问题，已经引起地球物理学家，特别是反演问题专家的严重关切。

非唯一性从何而来？反演理论之父——Buckus 和 Gilbert 认为：

“来源于观测数据的数目并非无限，以及观测数据具有误差。仅此而已！”

为了说明解的非唯一性，让我们来考虑一下零向量和零空间的概念。在模型可用离散参数表示的情况下，有(1.17)式，即

$$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}$$

这里的 $\boldsymbol{d}$ 和 $\boldsymbol{m}$ 分别为观测数据向量和模型参数向量，$\boldsymbol{G}$ 称为数据核。

假设上式数据方程有两个非零解 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$，即解是非唯一的，则有：

$$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}_1$$

和

$$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}_2$$

以上二式相减得

$$\boldsymbol{G}(\boldsymbol{m}_1-\boldsymbol{m}_2)=\boldsymbol{0} \tag{1.20}$$

由于两个解 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 是不同的，因而其差 $\boldsymbol{m}^{null}=\boldsymbol{m}_1-\boldsymbol{m}_2$ 不是零。但是 $\boldsymbol{G}(\boldsymbol{m}_1-\boldsymbol{m}_2)=\boldsymbol{0}$，所以 $\boldsymbol{m}^{null}$ 叫零向量，由零向量组成的空间叫零空间。可见，任何具有零空间的线性反演问题的解都是非唯一的。

如果，设 $\boldsymbol{m}^{par}$ 是数据方程 $\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}$ 的一个非零解，则 $\boldsymbol{m}^{par}+\alpha\boldsymbol{m}^{null}$ 也是一个非零解。这里 α 是一个不为零的任意常数。如果，某线性问题有 g 个独立的零向量，则其一般解为：

$$\boldsymbol{m}^{gen}=\boldsymbol{m}^{par}+\sum_{i=1}^{g}\alpha_i\boldsymbol{m}_i^{null} \tag{1.21}$$

式中：gen 代表一般解。

由(1.20)式可以看出，凡是和数据核 $\boldsymbol{G}$ 正交的模型空间中的向量均是零向量 $\boldsymbol{m}^{null}$。显然，对许多地球物理问题而言，$\boldsymbol{m}^{null}$ 不但存在，而且由零向量构成的零空间的范围是不小的。

与线性方程类似，如果模型参数是连续函数，则(1.14)式成立，即

$$d_j=\int_0^{\infty} g(z,\tau_j)m(z)\mathrm{d}z \qquad (j=1,2,\cdots,M)$$

式中：d_j 是第 j 个观测数据；$g(z,\tau_j)$是第 j 个核函数，$m(z)$为模型。上式也可以看成是一种线性变换，则有

$$d_j=\mathrm{L}_j[m(z)] \tag{1.22}$$

这里 L 是线性算子。

和线性方程一样，如果(1.14)式有两个非零解，比如说 $m_1(z)$和 $m_2(z)$，则

$$\int_0^{\infty} g(z,\tau_j)(m_1(z)-m_2(z))\mathrm{d}z$$

$$=\int_0^{\infty} g(z,\tau_j)m^{anah}(z)\mathrm{d}z \qquad (j=1,2,\cdots,M)$$

$$=\mathrm{L}_j[m^{anah}(z)]=0$$

式中：$m^{anah}(z)=m_1(z)-m_2(z)\neq 0$ 叫零化子。由零化子组成的空间叫零化空间。因此，任何具有零化空间的线性积分方程的反演问题，其解都是非唯一的。

换言之，非唯一性问题意味着拟合观测数据的模型分别位于两个独立的空间，即 m^{par}和 m^{null}(或 m^{anah})，之中，且有

$$\boldsymbol{d}=\boldsymbol{G}(\boldsymbol{m}^{par}+\boldsymbol{m}^{null})=\boldsymbol{G}\boldsymbol{m}^{par}+\boldsymbol{G}\boldsymbol{m}^{null}$$

和

$$d_j=\mathrm{L}_j[m^{par}+m^{anah}]=\mathrm{L}_j[m^{par}]+\mathrm{L}_j[m^{anah}] \qquad (j=1,2,\cdots,M)$$

模型空间 $\boldsymbol{m}^{par}$映射到数据空间与 $\boldsymbol{d}$ 相对应，而 m^{anah}的映射只与 $\boldsymbol{0}$ 相对应，如图 1.1 所示。

可见，反演一组地球物理观测数据，就是在模型空间中寻求一个特殊解，使之拟合观测数据。但是，如上所述，在这个特殊模型上加上任何一个零向量(或零化子)，所得到的模型均能拟合观测数据，所以解是非唯一的！

为了说明线性反演的非唯一性，让我们分析以下简单实例。

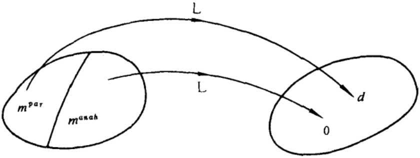


图 1.1 0空间示意图

例 3 设

$$d=\begin{bmatrix}\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\end{bmatrix}\begin{bmatrix}m_1\\m_2\\m_3\\m_4\end{bmatrix}$$

显然，其一个特解是 $\boldsymbol{m}^{par}=[d\ d\ d\ d]^{\mathrm{T}}$。而零向量可能是：

$$\boldsymbol{m}_1^{null}=\begin{bmatrix}1\\-1\\0\\0\end{bmatrix};\qquad \boldsymbol{m}_2^{null}=\begin{bmatrix}1\\0\\-1\\0\end{bmatrix};\qquad \boldsymbol{m}_3^{null}\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}$$

故其一般解为：

$$\boldsymbol{m}^{gen}=\begin{bmatrix}d\\d\\d\\d\end{bmatrix}+\alpha_1\begin{bmatrix}1\\-1\\0\\0\end{bmatrix}+\alpha_2\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}+\alpha_3\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}$$

式中：$\alpha_1,\alpha_2,\alpha_3$ 是不为零的参数。

十分清楚，线性反演中的非唯一性问题是十分重要又而必须认

真对待的问题。应该说，原则上人们对非唯一性是承认的，然而，对其重要性和严重程度却不是人人都认识到了的。

§6　结果的评价

既然，所有地球物理反演问题，包括线性反演和非线性反演问题，解都是非唯一的，只是程度不同而已。那么，所获得的解又有何价值呢？它是否包含了真实地球物理模型参数所具有的唯一信息呢？这就是结果评价所要研究的问题。

Backus 和 Gilbert 在其关于线性评价的文章中证明：在一定条件下，模型参数 $\boldsymbol{m}$ 在平均函数 A 的作用范围内的平均值 $\langle m\rangle=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}$，是能从观测数据 $\boldsymbol{d}$ 中所能求得的模型参数之唯一信息，而这种信息是不随解反演问题的方法以及所加的条件而改变的。换言之，如果真实的地球物理模型为 $\boldsymbol{m}'$，其对应的观测数据为 $\boldsymbol{d}$，而 $\boldsymbol{m}_1,\boldsymbol{m}_2,\cdots,\boldsymbol{m}_N$ 均为拟合观测数据 $\boldsymbol{d}$ 的解，Backus-Gilbert 的评价理论说明，在平均函数 $\boldsymbol{A}$ 的窗口范围内，下式成立，

$$\langle m\rangle=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}_1=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}_2=\cdots=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}_N=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}$$

因而，其平均值都是相等的，也是唯一能从观测数据中提取出来的有用信息。可见，任何能拟合观测数据的解，都是有意义的，因为它都包含了与真实地球物理模型相同的唯一信息。评价理论也进一步说明，反演工作者应该把注意力集中在平均函数窗口范围内的平均值 $\langle m\rangle$，而不应该仅关心在某一深度 z 处模型参数的数值。否则，就会得出错误的结论。由此可见，线性评价是对线性反演问题非唯一性进行估价的一种极其重要的方法，也是线性反演理论的重要组成部分。

同理，对连续模型而言，以下等式成立，

$$\begin{aligned}\langle m(z_0)\rangle&=\int_0^{\infty}A(z,z_0)m'(z)\mathrm{d}z=\int_0^{\infty}A(z,z_0)m_1(z)\mathrm{d}z\\&=\cdots=\int_0^{\infty}A(z,z_0)m_N(z)\mathrm{d}z\end{aligned}$$

式中：$A(z,z_0)$为平均函数；$m'(z)$代表真实地球物理模型；$m_1(z)$，…，$m_N(z)$代表能拟合观测数值的、通过反演所求得的模型。

不管对离散模型还是连续模型，$\langle m(z)\rangle$能在多大程度上反映深度 z 处地球物理参数的实际值，完全取决于平均函数 $A(z,z_0)$（或 $A(z,z_0)$）的形态。比如，$\boldsymbol{m}$ 代表速度时，且有

$$\boldsymbol{m}=[v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6 \quad v_7 \quad v_8]^{\mathrm{T}}$$

而平均函数

$$\boldsymbol{A}=\begin{bmatrix}0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0\end{bmatrix}^{\mathrm{T}}$$

则

$$\langle m\rangle=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}=\sum_{i=1}^{8}a_i v_i$$

显然，可以认为$\langle m\rangle$是 v_4 的模型参数的局部平均值。但是，如果

$$\boldsymbol{A}=\left[\frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8}\right]^{\mathrm{T}},$$

则$\langle m\rangle$就不再代表某一深度速度之局部平均值了。换句话说，该地球物理问题分辨力是很低的。

§7　解的稳定性

如前所述，反演问题就是从数据空间到模型空间的映射问题。如果数据空间有一个小范围的变化，相应于模型空间存在一个大范围的变化，如图 1-2 所示，则称这种映射或反演是不稳定的。

实践证明，地球物理学中的反演问题都是不稳定的。只是严重程度不同罢了。解不稳定，会造成迭代过程中的振荡，甚至会出现不收敛，给反演工作造成极大的麻烦。因此，提高解的稳定性，是反演中必须采取的重要措施。

为了说明在地球物理资料反演中解的不稳定性问题，让我们来研究一下在地球物理学中经常遇到的阿贝尔（Abel）方程的反演问题。其正演公式为：

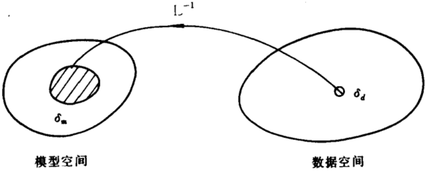


图 1-2　数据空间与模型空间的映射

$$d(x)=\alpha\int_x^{r_0}\frac{m(r)}{\sqrt{r^2-x^2}}\mathrm{d}r \tag{1.23}$$

显然，这是一个线性积分方程。如将 $d(x)$ 视为观测数据，$m(r)$ 视为模型，则 $g(r,x)=1/\sqrt{r^2-x^2}$ 就是核函数。显然根据观测数据 $d(x)$ 求模型 $m(r)$ 是一个典型的连续介质的反演问题。值得庆幸的是 (1.23)式中的 $m(r)$ 可以有解析表达式，即

$$m(r)=-\frac{1}{\pi}\int_r^{r_0}\frac{d'(x)}{\sqrt{x^2-r^2}}\mathrm{d}x \tag{1.24}$$

从(1.24)式可以看出，观测数据 $d(x)$ 的微小变化，必然影响 $d'(x)$，因而导致模型 $m(r)$ 发生很大的变化，出现反演的不稳定。下面，以解析的办法进行分析。

如果，$d(x)$ 有一观测误差 $\delta d(x)=a\cos kx$，则

$\delta d'(x)=-ak\sin kx$，代入(1.24)式，得：

$$\delta m(r)=\frac{1}{\pi}\int_r^{r_0}\frac{ak\sin kx}{\sqrt{x^2-r^2}}\mathrm{d}x$$

若 $r=0$，则

$$\delta m(0)=\frac{1}{\pi}\int_{0}^{x_0}\frac{\alpha k\sin kx}{x}\mathrm{d}x>0.452\ \alpha k$$

可见，只要 αk 足够大，那么观测数据 $d(x)$ 的微小扰动，将引起模型参数的巨大误差，导致反演结果变得极不稳定。

解的不稳定问题，许多作者又称为解的病态问题。病态的严重程度取决于数据核或核函数之间是否相关，相关说明稳定性低；不相关说明稳定性高。关于这个问题，我们将在讲述核函数性质时详细论述。

§8 线性反演问题综述

在这一节里，将从理论上进一步论述在线性反演问题中解的存在性，模型构制，非唯一性等问题。为使问题论述简单明了而又不失一般性，这里只限于一维问题。读者很容易把它推广到多维空间的地球物理反演问题。

在连续介质条件下，由(1.14)式知

$$d_j=\int_0^{\infty}g(z,\tau_j)m(z)\mathrm{d}z \qquad (j=1,2,\cdots,M)$$

且

$$m(z)\in \mathrm{L}_2(a,b) \tag{1.25}$$

在观测数据数目有限的情况下，上式可写为

$$\begin{aligned} d_j &=d(\tau_j)=(g(z,\tau_j),m(z)) \\ &=(g_j,m) \qquad (j=1,2,\cdots,M) \end{aligned} \tag{1.26}$$

式中：(g_j,m)表示内积。

在讨论线性反演理论若干问题之前，先对观测数据和核函数作如下假设：

(1)$d_j,j=1,2,\cdots,M$ 是没有误差的精确数据；

(2)$g_j,j=1,2,\cdots,M$ 是线性无关的。

若

$$d_j=(g_j,m) \qquad (j=1,2,\cdots,M)$$

满足上述两个假定条件。

首先，用这组线性无关的核函数 $g_j(z)$ 构成一组正交函数，有

$$\psi_k(z)=\sum_{j=1}^{M}\alpha_{kj}g_j(z) \qquad (k=1,2,\cdots,M) \tag{1.27}$$

显然，

$$(\psi_i,\psi_j)=\delta_{ij} \tag{1.28}$$

可见 Ψ 或 (g) 是无限维 Hilbert 空间中的 M 维向量：

$$\Psi=\begin{bmatrix}\psi_1\\ \psi_2\\ \vdots\\ \psi_M\end{bmatrix}; \qquad g=\begin{bmatrix}g_1\\ g_2\\ \vdots\\ g_M\end{bmatrix}$$

其次，以 α_{kj} 为系数，对观测数据 d_j 作一线性组合，并令其为 E_k，则有：

$$\begin{aligned}E_k&=\sum_{j=1}^{M}\alpha_{kj}d_j=\sum_{j=1}^{M}\alpha_{kj}(g_j,m)\\&=\left(\sum_{j=1}^{M}\alpha_{kj}g_j,m\right)\\&=(\psi_k(z),m)\end{aligned} \tag{1.29}$$

可见，E_k 是模型 $m(z)$ 在正交坐标基 $\psi_k(z)$ 轴上的投影。

第三，因为在区间 $[0,\infty]$ 上的任何函数 $m(z)$ 都可表示为级数，即

$$m(z)=\sum_{k=1}^{\infty}\beta_k\phi_k(z) \tag{1.30}$$

这里，$\phi_k(z)$ 可以视为 Hilbert 空间的任意正交坐标基。如果将上式分解为两项，并取

$$\left.\begin{aligned}&\phi_k(z)=\psi_k(z) && (k=1,2,\cdots,M)\\&\phi_k(z)=\text{任意坐标基} && (k>M)\end{aligned}\right\} \tag{1.31}$$

则

$$m(z)=\sum_{k=1}^{M}\beta_k\psi_k(z)+\sum_{k=M+1}^{\infty}\beta_k\phi_k(z) \tag{1.32}$$

可以证明 $\beta_k=E_k$.

因为

$$\begin{aligned}E_k&=(\psi_k(z),m)\\&=(\psi_k(z),\sum_{l=1}^{M}\beta_l\psi_l(z)+\sum_{l=M+1}^{\infty}\beta_l\phi_l(z))\\&=\left(\sum_{l=1}^{M}\beta_l(\psi_k(z),\psi_l(z))+\sum_{l=M+1}^{\infty}\beta_l(\psi_k(z),\phi_l(z))\right)\\&=\left(\sum_{l=1}^{M}\beta_l\delta_{kl}+0\right)=\beta_k\end{aligned}$$

所以，代入(1.32)式得，

$$\begin{aligned}m(z)&=\sum_{k=1}^{M}E_k\psi_k(z)+\sum_{k=M+1}^{\infty}\beta_k\phi_k(z)\\&=\sum_{k=1}^{M}E_k\psi_k(z)+m^{anah}(z)\end{aligned} \tag{1.33}$$

讨　论：

(1) 给定一组观测数据：

$d_j=(g_j,m)\qquad(j=1,2,\cdots,M)$

由上面的分析可以看出，只要满足两个假定条件，总可以找到一个为(1.33)式所示的能拟合观测数据的模型 $m(z)$ 与之对应。换言之，解是存在的。

(2) 根据观测数据所构制的模型 $m(z)$ 始终由两部分组成。第一部分为：

$$\sum_{k=1}^{M}E_k\psi_k(z)$$

它取决于观测数据 $d_j=(g_j,m(z))$。

第二部分为：

$$m^{anah}(z)=\sum_{k=M+1}^{\infty}\beta_k\phi_k(z)$$

它与观测数据无关。由(1.33)式可知，模型的构制过程，本质上就是对线性无关的核函数 $g_j(z)$ 实行正交变换，求得相应的新正交坐标基 $\psi_k(k=1,2,\cdots,M)$ 及模型在这个正交坐标基上投影的过程。

(3) 式(1.33)表明，反演问题的解是非唯一的，因为反演构制出的模型由两部分所组成。第一部分是特殊解，它与 M 个观测数据有关，为 M 维；第二部分是零空间中的零化子向量，与观测数据无关，有∞维。在特殊解上加以任何零化子向量所得到的模型，都可拟合观测数据，所以解是非唯一的。

将(1.33)代入数据方程(1.26)式，则有：

$$\begin{aligned}
d_j&=(g_j,m(z))=(g_j,\sum_{k=1}^{M}E_k\psi_k(z)+\sum_{k=M+1}^{\infty}\beta_k\phi_k(z))\\
&=(g_j,\sum_{k=1}^{M}E_k\psi_k(z))+(g_j,\sum_{k=M+1}^{\infty}\beta_k\phi_k(z))\\
&=\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}d_l\sum_{i=1}^{M}\alpha_{ki}(g_i,g_j)+\sum_{k=M+1}^{\infty}\beta_k(g_j,\phi_k(z))\\
&=\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}d_l\alpha_{kj}+\sum_{k=M+1}^{\infty}\beta_k(g_i,\phi_k(z))\\
&=\sum_{k=1}^{M}\alpha_{kj}^2d_j+\sum_{k=M+1}^{\infty}\beta_k(g_j,\phi_k(z))\\
&=d_j \qquad (j=1,2,\cdots,M) \qquad (1.34)
\end{aligned}$$

这里，我们利用了(1.27)、(1.29)、(1.32)等式。(1.34)式说明，在由 M 维观测数据决定的 M 维特殊解

$$\sum_{k=1}^{M}E_k\psi_k(z)$$

上加以零化向量 $m^{anah}(z)$ 模型，仍能拟合观测数据。这就无可争辩地说明，线性反演的解是非唯一的，即解的非唯一性是存在的。

(4) 在所有能拟合观测数据的模型中，L_2 范数为最小的模型，

必然是

$$\| m^{anah}(z) \|^2 = 0$$

的模型，即“最小模型”。因为

$$\| m(z) \|^2 = \left\| \sum_{k=1}^{M} E_k \psi_k(z) + m^{anah}(z) \right\|^2$$

$$\leqslant \left\| \sum_{k=1}^{M} E_k \psi_k(z) \right\|^2 + \| m^{anah}(z) \|^2$$

$$= \sum_{k=1}^{M} E_k^2 + \| m^{anah}(z) \|^2$$

欲使 $\| m(z) \|^2$ 为最小，必须 $\| m^{anah}(z) \|^2 = 0$。显然，最小模型是能拟合观测数据，而又无零化空间影响的那个模型。最小模型是正交坐标系 $\psi_k(z)(k=1,2,\cdots,M)$ 中的一个 M 维向量，可以把它看成是核函数 $g_j(z)$ 的一种线性组合，即

$$m(z) = \sum_{k=1}^{M} E_k \psi_k(z) = \sum_{j=1}^{M} r_j g_j(z) \qquad (j=1,2,\cdots,M) \quad (1.35)$$

总之，根据观测数据假设 $\| m(z) \|^2$ 为最小求得的模型是最小模型；最小模型是核函数的线性组合；而模型的构制过程，实际上就是寻找正交坐标基 ψ_k 的过程。

第二章　参数化模型的最小长度解

在第一章中已经说明，所谓参数化模型是指那些能用有限个参数表征的模型，既包括能用有限参数表征的一维、二维、三维模型，也包括能用有限参数表征的离散模型和连续模型。简言之，参数化模型是指参数为有限的那些模型。

在正式论述本章内容之前，我们还必须弄清最小长度的含义。在处理超定问题时，为处理观测数据中所提供的“多余”信息而采用的误差向量为最小（即方差最小）的准则；在解欠定问题时，为处理观测数据所提供的信息不是以确定全部模型参数，而采用模型参数向量的长度为最小的准则。这里误差向量为最小，或模型参数向量长度为最小（都是欧基里德空间的长度），是 L_2 范数意义下的长度，或范数为最小。长度的定义不同，所求得的模型也会发生变化。在地球物理资料反演问题中，经常应用 L_2 范数，L_1 范数和 L_∞ 范数等作为测量长度的定义。基于这种概念所求得的解，统称为长度解。

§1　线性反演问题的最小方差解

在线性反演问题中，如果观测数据的个数多于模型参数的个数，更准确地说，在 $M>N=r$ 的情况下，最简单、最常用的反演方法是最小方差法。这里 r 是数据方程

$$\underset{(M\times1)}{\boldsymbol{d}} = \underset{(M\times N)}{\boldsymbol{G}}\ \underset{(N\times1)}{\boldsymbol{m}} \tag{2.1}$$

中数据核 $\boldsymbol{G}$ 的秩。

设 $\boldsymbol{e}$ 为观测数据 $\boldsymbol{d}$ 与理论计算值 $\boldsymbol{Gm}$ 之误差向量，则方差（即目标函数）为

$$\boldsymbol{E}=\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}=(\boldsymbol{d}-\boldsymbol{Gm})^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{Gm}) \tag{2.2}$$

式中：

$$
e=\begin{bmatrix} d_1- & \sum_{i=1}^{M} G_{1i}m_i \\ d_2- & \sum_{i=1}^{M} G_{2i}m_i \\ \vdots & \vdots \quad \vdots \\ d_m= & \sum_{i=1}^{M} G_{mi}m_i \end{bmatrix} \tag{2.3}
$$

将(2.2)式展开得：

$$
\begin{aligned} E &= (\boldsymbol{d}^{\mathrm{T}}-\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}})(\boldsymbol{d}-\boldsymbol{Gm}) \\ &= \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}-\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}-\boldsymbol{d}^{\mathrm{T}}\boldsymbol{Gm}+\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{Gm} \end{aligned}
$$

最小方差解必须满足：

$$
\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{m}^{\mathrm{T}}}=-\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}+\boldsymbol{G}^{\mathrm{T}}\boldsymbol{Gm}=0 \tag{2.4}
$$

所以 $\boldsymbol{m}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}$

或对 m 求偏导数，并令其为 0，则有

$$
\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{m}}=-\boldsymbol{d}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}=0
$$

同理，可得如式(2.4)所示正态方程。

例：一组观测数据为：

d_j,　　$(j=1,2,\cdots,M)$　　$(M>5)$

现欲用一个平面方程

$d_j=m_1+m_2x_j+m_3y_j$

拟合之。

式中：m_1,m_2,m_3 为模型参数；x_j 和 y_j 是第 j 个数据对应的坐标。由(2.1)知，数据方程如下：

$$
\boldsymbol{d}=\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{bmatrix}=\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_M & y_M \end{bmatrix}\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_3 \end{bmatrix}=\boldsymbol{Gm}
$$

所以

$$
\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_M \\ y_1 & y_2 & \cdots & y_M \end{bmatrix}\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_M & y_M \end{bmatrix}
$$

$$
=\begin{bmatrix} M & \Sigma x_j & \Sigma y_j \\ \Sigma x_j & \Sigma x_j^2 & \Sigma y_j x_j \\ \Sigma y_j & \Sigma x_j y_j & \Sigma {y_j}^2 \end{bmatrix}
$$

而

$$
\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_M \\ y_1 & y_2 & \cdots & y_M \end{bmatrix}\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{bmatrix}=\begin{bmatrix} \Sigma d_j \\ \Sigma x_j d_j \\ \Sigma y_j d_j \end{bmatrix}
$$

故最小方差解为

$$
\boldsymbol{m}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}
$$

$$
=\begin{bmatrix} M & \Sigma x_j & \Sigma y_j \\ \Sigma x_j & \Sigma x_j^2 & \Sigma y_j x_j \\ \Sigma y_j & \Sigma x_j y_j & \Sigma {y_j}^2 \end{bmatrix}^{-1}\begin{bmatrix} \Sigma d_j \\ \Sigma x_j d_j \\ \Sigma y_j d_j \end{bmatrix} \tag{2.5}
$$

讨　论：

(1) 从线性代数基本理论可知，在线性方程

$$
\boldsymbol{G}^{\mathrm{T}}\boldsymbol{Gm}=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}
$$

中，如果系数矩阵 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}(N\times N)$ 的秩 $r<N$，则方程是奇异的，即使 M

$>N$,此时在 M 个观测数据中,没有足够能确定 N 个未知数的观测数据,或者说,观测数据没有提供确定 N 个模型参数的独立的信息。方程是奇异的,意味着 G^TG 或 G 是奇异的,奇异矩阵是无常规意义下的解。这种情况,在地球物理资料的反演中是经常遇到的。

(2) G^TG 有零特征值存在时,数据方程(2.1)式是奇异的,无法求解;当 G^TG 的特征值很小时,方程(2.1)式是病态的,会使解变得极不稳定,使反演非常困难,甚至无法收敛。这也是在地球物理资料反演中经常遇到的问题。因此,克服反演过程中的奇异和病态问题,就成了反演理论必须解决的重要课题。

§2 纯欠定问题的解法

所谓纯欠定问题,是指在(2.1)式中,未知参数的个数 N 大于观测数据的个数 M,且矩阵 G 的秩 $r=M$ 的情况。换言之,在 M 个方程中,既无相关方程,也无矛盾方程存在。从线性代数理论可知,此时有无限多个解能满足线性方程组(2.1)式,且其误差均为零。这是因为,虽然观测数据提供了一些确定模型参数的信息,但其数量不足以全部确定模型参数,或未提供确定模型参数的足够充分的信息。因此,解不是唯一的,甚至有无限多能拟合观测数据的解。

为了求得反演问题的一个解,我们必须从无限多个能拟合观测数据的解中,挑选出一个我们所需要的特定解。因此,解方程(2.1)式时,必须加上一些它未包含的信息。这种附加给反演问题的信息叫"先验信息"(priori information)。

应说明,加先验信息的目的是为了补充那些为确定模型参数所不足的信息。因此,为了使反演问题的解更切合实际情况,就应本着"缺什么信息,就补充什么信息"的原则。为了更好地运用"择缺补充"的原则,先分析一下有哪几类可能的先验信息。

第一类先验信息是待求地球物理参数的物理性质和其可能的数值范围。如速度(v),密度(σ)和电阻率(ρ)等的非负性,它们都不可能

小于零。而且它们的数值，根据一般物理常识可以限制在一定范围以内，比如，200 m/s$\leqslant v \leqslant$8 000 m/s，0.0 g/cm$^3 \leqslant \sigma \leqslant$16.0 g/cm^3 等等。

第二类先验信息来自于其他已知的地质、地球物理和钻井资料。比如反演地区基底的埋深，油层的厚度，或金属矿的属性等等。

第三类，某些参数比其他参数对解决地球物理问题更重要，此时可以对模型参数进行加权，在一定权系数约束下求解。

第四类，也是纯欠定问题解法中常应用的先验信息，假定的地球物理模型"最简单"。这里所谓最简单是指在保留实际地球物理模型基本特征不变的情况下，对地球物理模型的一种简化。解的长度，比如说解的欧基里德长度为最小的模型，应该是一种最简单的模型。当然，这里定义的"简单"并不一定处处都非常合理。因此，又出来了其他形式的最简单模型，如该模型的变化为最小的 L_2 范数意义下的模型等。由 L_2 范数定义的 $E=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}=\sum_{i=1}^{N} m_j{}^2=\min$ 的最简单模型，有其天然的合理性，因为它保留了模型所具有的基本特征。

除此之外，在实际反演过程中，还可能遇到其他各种先验信息，这里就不一一列举了。

设(2.1)式是一纯欠定问题，此时的目标函数，在(2.1)式约束之下有极小，即

$$E=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}$$

根据极值理论，必须引入拉格朗日算子"$\boldsymbol{\lambda}$"将条件极值问题化为无条件极值问题。因此，目标函数应为：

$$E=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}+\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}) \tag{2.6}$$

显然，求上述目标函数的极小值问题，可以化为求：

$$\frac{\partial E}{\partial \boldsymbol{m}}=\boldsymbol{m}^{\mathrm{T}}-\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{G}=0$$

之解，故

$$m = G^T \lambda \tag{2.7}$$

将(2.7)代入(2.1)，则

$$d = GG^T \lambda \tag{2.8}$$

故

$$\lambda = (GG^T)^{-1} d \tag{2.9}$$

再将(2.9)代入(2.7)，得纯欠定问题的正态方程

$$m = G^T((GG^T)^{-1} d \tag{2.10}$$

例：在三维空间(x、y、z)坐标中，过原点求一向量，使其在(x、y)平面上的坐标为(x_0,y_0)。

十分明显，这是一个纯欠定问题。因为，在通过(x_0,y_0)垂直于(x,y)平面划的一条垂线上的每一点在(x,y)平面上的坐标都是(x_0,y_0)，所述问题有无限多个解。在这无限个解中，只有 $z=0$ 这一个解才满足长度为最小，且在(x,y)平面上的坐标为在(x_0,y_0)这一条件。

欠定问题在地球物理资料的反演中是经常遇到的。如果比较一下(2.10)式和(2.4)式，会发现在最小方差解的超定问题中，须求对称矩阵 G^TG 的逆，这里矩阵为 $N \times N$ 阶方阵。而在解纯欠定问题时，须求对称矩阵 GG^T 的逆，它的阶数为 $M \times M$ 阶方阵。和超定问题一样，解欠定问题时，也存在方程(2.1)式的"奇异"和"病态"两种棘手问题。所谓奇异问题，是指 GG^T 的特征值中有为零的问题；病态问题，是指 GG^T 中有小特征值的问题。这都是在解欠定问题时必须认真对待的。

§3 混定问题的解法——马夸特(Marquardt)法

当线性反演问题

$$\underset{(M\times 1)}{d} = \underset{(M\times N)}{G} \ \underset{(N\times 1)}{m}$$

呈现 $\min(M,N) > r$ 的情况时，称为混定问题。解混定问题的方法，通常称马夸特法，或脊回归法(ridge regression)，又称为阻尼最小二

乘法(damping R.M.S)。

从 M、N 和 r 的关系看，由于矩阵 $\boldsymbol{G}$ 的秩 r 意味着在方程(2.1)式中，只有 r 个线性无关的方程，只能确定 r 个非零的解。因此，$M>r$ 是超定问题，而 $N>r$ 又是欠定问题。许多地球物理线性反演问题既不完全是超定问题，也不完全是欠定问题，常常表现为一种混定形式(即混定问题)。

鉴于混定问题的特殊性，它既有超定问题，也有欠定问题的性质，因此不难设想其目标函数应兼有方差项 $(\boldsymbol{d}-\boldsymbol{Gm})^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{Gm})$ 和模型长度项 $\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}$ 两项内容，即

$$
\begin{aligned}
E &= (\boldsymbol{d}-\boldsymbol{Gm})^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{Gm})+\in^{2}\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m} \\
&= \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}-\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}-\boldsymbol{d}^{\mathrm{T}}\boldsymbol{Gm}+\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{Gm}+\in^{2}\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}
\end{aligned}
\tag{2.11}
$$

求 E 相对于 $\boldsymbol{m}$ 或 $\boldsymbol{m}^{\mathrm{T}}$ 的偏导数，并设其为零，简化后得

$$
[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\in^{2}\boldsymbol{I}]\boldsymbol{m}=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d} \tag{2.12}
$$

因而

$$
\boldsymbol{m}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\in^{2}\boldsymbol{I})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d} \tag{2.13}
$$

式中：$\in^{2}$ 称阻尼系数或加权因子，它决定预测误差项和模型 L_2 范数长度项在极小化目标函数 E 时之相对重要性。如果 $\in^{2}$ 足够大，则模型的 L_2 范数长度在极小化过程中起着主要作用，或者说是极小解的欠定部分；如果 $\in^{2}$ 为零，则极小的是方差部分，或者说解的超定部分。然而，对大多数地球物理反演问题而言，$\in^{2}$ 为零或足够大都难以取得模型的最优解。要寻找一种折衷方案，找到一个合适的 $\in^{2}$ 值，这就要在迭代过程中不断修改 $\in^{2}$ 的大小。这里不存在一种简单的计算最佳阻尼系数 $\in^{2}$ 的方法，只能在反演过程中用“尝试法”确定。由此可见，调整 $\in^{2}$ 的大小，本质上就是寻找在迭代过程中的最佳校正方向和最佳校正步长！

根据对称矩阵 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 的正交分解，$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 可以分解为：

$$
\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}=\boldsymbol{R}\Lambda\boldsymbol{R}^{\mathrm{T}} \tag{2.14}
$$

这里 Λ 是 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 之特征值构成的对角线矩阵，而 $\boldsymbol{R}$ 是 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 之特征向

量矩阵，且满足

$$R^{T}R=RR^{T}=I$$

所以，(2.10)式中的系数矩阵可写为

$$[G^{T}G+\in^{2}I]=[R\Lambda R^{T}+\in^{2}I]=R\Lambda' R^{T} \tag{2.15}$$

式中：

$$\Lambda'=\begin{bmatrix} \lambda_1+\in^2 & & & \\ & \lambda_2+\in^2 & & 0 \\ 0 & & \ddots & \\ & & & \lambda_r+\in^2 \end{bmatrix} \tag{2.16}$$

且 λ_i 是对称矩阵 $G^{T}G$ 之第 i 个特征值。

十分明显，由于对角线矩阵 Λ' 是在 Λ 的各对角线要素上加了一个正数 $\in^2$，从而极大地改善了系数矩阵 $[G^{T}G+\in^{2}I]$ 相对于 $[G^{T}G]$ 而言的求逆条件。此时的条件数为：

$$\mathrm{cond}[G^{T}G+\in^{2}I]=\frac{\lambda^{\max}+\in^{2}}{\lambda^{\min}+\in^{2}}<\frac{\lambda^{\max}}{\lambda^{\min}}=\mathrm{cond}[G^{T}G],$$

马夸特法的系数矩阵有较好的条件数，对地球物理资料反演十分有利。在数据方程中 G 出现奇异或病态时，马夸特法往往可以稳定求解。正因为如此，阻尼最小二乘法于过去一段时间里在地球物理资料的反演中，已被广泛采用，并取得了世人瞩目的成果。

这里，还应再一次指出，马夸特法应用的条件是数据方程呈混定的时候，也就是应用于其系数矩阵 G 为奇异和病态的时候！

§4 先验信息在模型构制中的应用

在本章§2中简要地叙述了在解欠定问题时，需要对模型参数强加一些先验信息，以限制 $(N-M)$ 个不足的信息。然而，这并不表明，在解超定问题时，就不应该也不可能对模型参数强加任何先验信息。恰恰相反，在解超定问题时，可对模型参数强加先验信息，在解欠

定问题时，也可对观测数据强加已知的先验信息。

1. 对模型参数的限制

在有些情况下，首先 $E=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}$ 为最小作为模型长度的定义并不完全适合。例如，解决岩石物性在空间的起伏变化时，并不希望所求的岩石物性在 E 接近零这种意义下的最小，而要求它为接近于其平均值这种意义下为最小。此时的长度定义为：

$$E=(\boldsymbol{m}-\langle\boldsymbol{m}\rangle)^{\mathrm{T}}(\boldsymbol{m}-\langle\boldsymbol{m}\rangle) \tag{2.17}$$

这里平均值或数字期望 $\langle\boldsymbol{m}\rangle$ 就是另一种形式的先验信息。

再者，令 $E=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}$ 为最小作为求取模型参数的测度，有时也不完全合理。有的学者认为 $E=\boldsymbol{m}'^{\mathrm{T}}\boldsymbol{m}'$ 或 $E=\boldsymbol{m}''^{\mathrm{T}}\boldsymbol{m}''$ 为最小作为求取模型参数的测度更为合理。这里“′”、“″”表示模型相对于深度的一阶、二阶偏导数。

第三，也可以认为某些模型参数重要，而另一些模型参数不重要，因而引入加权因子的概念，定义一种新的测度：

$$E=[\boldsymbol{Dm}]^{\mathrm{T}}[\boldsymbol{Dm}]=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{Dm}=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{W}_m\boldsymbol{m} \tag{2.18}$$

式中：矩阵 $\boldsymbol{W}_m=\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}$ 就是加权因子。式(2.18)定义的是一种加权测度，此时的欠定问题的最小长度解为：

$$\boldsymbol{m}=\boldsymbol{W}_m\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{W}_m\boldsymbol{G}^{\mathrm{T}}]^{-1}\boldsymbol{d}$$

与(2.17)式对应的欠定问题的最小长度解为：

$$\boldsymbol{m}=\langle\boldsymbol{m}\rangle+\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}[\boldsymbol{d}-\boldsymbol{G}\langle\boldsymbol{m}\rangle] \tag{2.19}$$

如将(2.17)式修改为

$$E=[\boldsymbol{m}-\langle\boldsymbol{m}\rangle]^{\mathrm{T}}\boldsymbol{W}_m[\boldsymbol{m}-\langle\boldsymbol{m}\rangle]$$

则(2.19)相应修改为：

$$\boldsymbol{m}=\langle\boldsymbol{m}\rangle+\boldsymbol{W}_m\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{W}_m\boldsymbol{G}^{\mathrm{T}}]^{-1}[\boldsymbol{d}-\boldsymbol{G}\langle\boldsymbol{m}\rangle] \tag{2.20}$$

2. 对观测数据的限制

如 $\boldsymbol{d}=\boldsymbol{Gm}$ 是一个超定系统。在解超定问题时，可以根据观测数据的精度或重要程度加以相应的权，则目标函数为

$$E=[\boldsymbol{d}-\boldsymbol{Gm}]^{\mathrm{T}}\boldsymbol{W}_e[\boldsymbol{d}-\boldsymbol{Gm}] \tag{2.21}$$

其最小方差解为：

$$m=[G^{T}W_{e}G]^{-1}G^{T}W_{e}d \qquad (2.22)$$

式中 W_e 为观测数据的加权因子。

对于混定问题，如极小目标函数

$$E=[d-Gm]^{T}W_{e}[d-Gm]+\in^{2}m^{T}W_{m}m \qquad (2.33)$$

则其解为：

$$m=[G^{T}W_{e}G+\in^{2}W_{m}]^{-1}G^{T}W_{e}d \qquad (2.24)$$

如极小为：

$$E=[d-G(m-\langle m\rangle)]^{T}W_{e}[d-G(m-\langle m\rangle)]$$
$$+\in^{2}[m-\langle m\rangle]^{T}W_{m}[m-\langle m\rangle] \qquad (2.25)$$

则其解为：

$$m=\langle m\rangle+W_{m}G^{T}[GW_{m}G^{T}+\in^{2}W_{e}]^{-1}[d-G\langle m\rangle] \qquad (2.26)$$

或

$$m=\langle m\rangle+[G^{T}W_{e}G+\in^{2}W_{m}]^{-1}G^{T}W_{e}[d-G\langle m\rangle] \qquad (2.27)$$

由此可见，在构制模型时，既可对观测误差加权，也可以对模型参数加权。加权就是强加一些已知的先验信息。先验信息不同，权系数变化，所得到的模型自然会有差异。但是，它们有一个共同的特点：都能拟合观测数据。这就说明，同一组地球物理观测数据，有许多可以拟合这组观测数据的模型与之对应。这就是解的非唯一性。

3. 等式限制条件的应用

在地球物理资料的反演过程中，有时还需要加一些等式限制条件。比如，在重力资料反演时，地表的密度是已知的；在地震资料反演中，某一地层中的波速是已知的等等，这是另一种常用的先验信息。现在来研究在解最小方差反问题时，如何增加这种等式限制条件。

设这类等式限制条件可以归纳为如下线性方程组：

$$Fm=h \qquad (2.28)$$

式中：F 是系数矩阵，为 $R\times N$ 阶

和§2中讨论的方法相同，可用拉格朗日算符法，将条件极值问

题化为无条件极值问题，即化为求目标函数，即

$$E=[\boldsymbol{d}-\boldsymbol{Gm}]^{\mathrm{T}}[\boldsymbol{d}-\boldsymbol{Gm}]+\lambda^{\mathrm{T}}[\boldsymbol{Fm}-\boldsymbol{h}] \tag{2.29}$$

的极值问题。设：

$$\frac{\partial E}{\partial \boldsymbol{m}}=-\boldsymbol{d}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{m}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\lambda^{\mathrm{T}}\boldsymbol{F}=0 \qquad \text{或}$$

$$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}-\boldsymbol{G}^{\mathrm{T}}\boldsymbol{Gm}-\boldsymbol{F}^{\mathrm{T}}\boldsymbol{\lambda}=0 \tag{2.30}$$

将(2.30)与(2.28)联立求解，则有

$$\begin{bmatrix}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} & \boldsymbol{F}^{\mathrm{T}}\\ \boldsymbol{F} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{m}\\ \boldsymbol{\lambda}\end{bmatrix}=\begin{bmatrix}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}\\ \boldsymbol{h}\end{bmatrix} \tag{2.31}$$

解(2.31)，即可求得 $\boldsymbol{m}$。

当然，也可以用其他方法增加这类等式限制条件。例如，将这个等式限制条件(2.28)写入线性方程组 $\boldsymbol{d}=\boldsymbol{Gm}$ 之中，并构成一个新的线性方程组

$$\boldsymbol{G}'\boldsymbol{m}=\boldsymbol{d}' \tag{2.32}$$

其中

$$\boldsymbol{G}'=\begin{bmatrix}\boldsymbol{G}\\ \boldsymbol{F}\end{bmatrix};\qquad \boldsymbol{d}'=\begin{bmatrix}\boldsymbol{d}\\ \boldsymbol{h}\end{bmatrix}$$

解此方程组即可求得模型参数 $\boldsymbol{m}$，显然它是满足(2.28)式的。

§5　观测数据和模型参数估算值之方差

前面讨论中，不管是超定问题，欠定问题还是混定问题，均未涉及观测数据的统计特征。实际地球物理资料的反演中，观测数据是有误差的，有误差就要遵守一定的统计特性。在欧基里德空间解地球物理反问题时，对观测数据的统计特性有何要求，这是要讨论的问题。

假定每一个观测数据都是随机变量，且服从高斯分布规律，即

$$p(d)=\frac{1}{\sqrt{2}\,\sigma}\exp\left[\frac{-(d-\langle d\rangle)^2}{2\sigma^2}\right] \tag{2.33}$$

式中：σ 为方差；$\langle d\rangle$ 为均值；$p(d)$ 表示随机变量 d 的概率分布函数。

对 M 个独立观测数据来说，其联合分布满足

$$p(d)=2^{-\frac{M}{2}}\sigma^{-M}\exp\left[-\frac{\sigma^{-2}}{2}\sum_{j=1}^{M}(d_j-\langle d\rangle)^2\right] \tag{2.34}$$

欲使随机变量 d 的概率最大，则必须使

$$E=\sum_{j=1}^{M}(d_j-\langle d\rangle)^2=\min$$

此即最小方差的目标函数。

可见，因 L_2 范数就意味着观测数据（或模型参数）必须遵守高斯正态分布的统计规律。实践证明，大多数地球物理观测数据都服从或近似服从高斯分布，这就为利用 L_2 范数极小求解地球物理问题提供了可靠的依据。

假定观测数据服从高斯分布，具零平均值，在方差为 σ_d 的条件下，分析一下在反演映射过程中，观测数据的误差对模型参数有何影响。

因为

$$\boldsymbol{d}+\boldsymbol{\sigma_d}=\boldsymbol{Gm}+\boldsymbol{G\sigma_m} \tag{2.35}$$

式中：

$$\boldsymbol{\sigma_d}=\begin{bmatrix}\sigma_{d1}\\ \sigma_{d2}\\ \vdots\\ \sigma_{dM}\end{bmatrix};\qquad \boldsymbol{\sigma_m}=\begin{bmatrix}\sigma_{m1}\\ \sigma_{m2}\\ \vdots\\ \sigma_{mN}\end{bmatrix}$$

式中：σ_d 为观测数据；σ_m 为模型参数之方差向量；σ_{dj}为第 j 个观测数据；而 σ_{mj}为第 j 个模型参数的方差。

利用数据方程(2.1)式，则(2.35)式可化为：

$$\boldsymbol{\sigma_d}=\boldsymbol{G\sigma_m} \tag{2.36}$$

故在最小方差意义下，有

$$\boldsymbol{\sigma_m}=[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\sigma_d} \tag{2.37}$$

由此可知，最小方差解 $\boldsymbol{m}=[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}d$ 之协方差矩阵为：

$$[\mathrm{cov}\boldsymbol{m}]=\sigma_m\sigma_m{}^{\mathrm{T}}=[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}[\mathrm{cov}\boldsymbol{d}][(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}]^{\mathrm{T}} \tag{2.38}$$

在模型最小长度解的意义下：

$$\sigma_m=\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}\sigma_d \tag{2.39}$$

其解(2.10)式的协方差矩阵为：

$$[\mathrm{cov}\boldsymbol{m}]=\sigma_m\sigma_m{}^{\mathrm{T}}=[\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}]\mathrm{cov}\boldsymbol{d}[\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}]^{\mathrm{T}} \tag{2.40}$$

式中：

$$[\mathrm{cov}\boldsymbol{d}]=\sigma_d\sigma_d{}^{\mathrm{T}} \tag{2.41}$$

如果说，观测数据是相互独立的，且均为单位标准方差，则

$$[\mathrm{cov}\boldsymbol{d}]=\sigma_d{}^2\boldsymbol{I} \tag{2.42}$$

因而，(2.38)式和(2.40)式分别简化为

$$\begin{aligned}[\mathrm{cov}\boldsymbol{m}]&=[[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}]\sigma_d{}^2\boldsymbol{I}[[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}]^{\mathrm{T}}\\&=\sigma_d{}^2[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\end{aligned} \tag{2.43}$$

或

$$\begin{aligned}[\mathrm{cov}\boldsymbol{m}]&=[\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}]\sigma_d{}^2\boldsymbol{I}[\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}]^{\mathrm{T}}\\&=\sigma_d{}^2\boldsymbol{G}^{\mathrm{T}}[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-2}\boldsymbol{G}\end{aligned} \tag{2.44}$$

从(2.43)和(2.44)式可以看出，不管是最小方差解还是模型最小的长度解，模型估计值之方差主要都取决于矩阵$[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]$，或$[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]$之特征值。特征值越小，引起的方差越大。由此可见，小特征值对模型参数的方差起着决定性的作用。

然而，小特征值对观测数据的误差影响不大(这一点将在下一章讨论)。在有小特征值时，观测数据的微小变化，将导致模型参数的巨大变化。小特征值对模型参数的方差贡献大，大特征值对观测数据的误差贡献大。这一重要结论说明：误差曲线的特征反映了$[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]$(或$[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]$)特征值的性质。如图 2.1(a)所示，其误差曲线在极小点附近非常尖锐，说明观测数据变化不会导致模型参数的明显变化，$[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]$的特征值差别不大，或者是无小特征值存在，所确定的模型参数比较精确；而图 2.1(b)所示正好相反，误差曲线平缓说明$[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]$的特征值差别较大，观测数据的小误差，可以导致模型参数的明显变化，因而

所确定模型的参数精度不高。

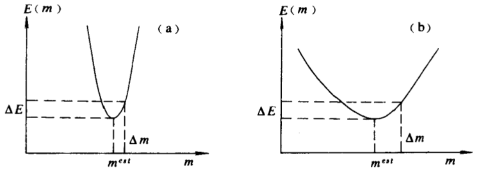


图 2.1 误差曲线的特征

因为误差函数曲线 $E(\boldsymbol{m})-\boldsymbol{m}$ 在极小点处的曲率是该点曲线尖锐程度的量度，所以，曲率必然和解的方差有着密切的关系。由极值定理可知，在极小点处 $\partial E(\boldsymbol{m})/\partial \boldsymbol{m}=0$，因此，在极小点附近，$E(\boldsymbol{m})$的台劳级数展开式可写为：

$$\begin{aligned}\Delta E(\boldsymbol{m}) &= E(\boldsymbol{m})-E(\boldsymbol{m}^0)\\ &\approx[\boldsymbol{m}-\boldsymbol{m}^0]^{\mathrm{T}}\boldsymbol{Q}[\boldsymbol{m}-\boldsymbol{m}^0]\end{aligned}\tag{2.45}$$

式中：$\boldsymbol{Q}$ 为(1.12)式所示，是赫米特矩阵；$\boldsymbol{m}^0$ 是初始模型。

又因为

$$E(\boldsymbol{m})=[\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}]^{\mathrm{T}}[\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}]$$

故

$$\boldsymbol{Q}=\frac{1}{2}\frac{\partial}{\partial \boldsymbol{m}}\left(\frac{\partial E(\boldsymbol{m})}{\partial \boldsymbol{m}^{\mathrm{T}}}\right)=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\tag{2.46}$$

将(2.46)代入(2.43)式，则在观测数据不相关且具单位方差的情况下，最小方差解之协方差矩阵为：

$$[\mathrm{cov}\boldsymbol{m}]=\sigma_d^2[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}=\sigma_d^2\boldsymbol{Q}^{-1}\tag{2.47}$$

由于在极小点处，函数曲线 $E(\boldsymbol{m})$-$\boldsymbol{m}$ 的曲率 $k=\boldsymbol{Q}$，这就说明模

型参数 $\boldsymbol{m}$ 的方差既可用观测数据的方差乘以从数据到模型之映射误差$[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}$，也可以用观测数据的方差除以该极小点处误差函数之曲率的公式计算出。

§6 线性规划——L_1 范数解

前面已经论述了，在观测数据的误差服从正态分布时，从统计学的观点看，此时应用 L_2 范数解是最合理的。如果观测数据 $\boldsymbol{d}$ 服从指数分布规律，即当

$$p(d)=\frac{1}{\sigma}\exp\left[\frac{-|d-\langle d\rangle|}{\sigma}\right] \tag{2.47}$$

时，又应该利用何种范数解线性方程(2.1)式呢？

如果对 M 个独立的观测数据来说，这时联合分布应为：

$$p(d)=2^{-\frac{M}{2}}\sigma^{-M}\exp\left[-\frac{1}{\sigma}\sum_{j=1}^{M}|d_j-\langle d\rangle|\right] \tag{2.48}$$

式中：$\langle d\rangle$为均值；σ 为方差；$p(d)$是观测数据的概率密度函数。

由(2.48)式知，欲使 $p(d)$最大，必须有：

$$E=\sum_{j=1}^{M}|d_j-\langle d\rangle|=\min$$

由此可见，当观测数据是随机变量，且服从指数分布时，应该用 L_1 范数解(2.1)式才是符合统计规律的。下面介绍的线性规划法是一种解 L_1 范数的行之有效的方法。

线性规划(Linear Programming)简称 LP，是一种求条件极值的方法。其目标函数和约束条件都是关于自变量的线性方程。所以，线性规划问题是求目标函数。

$$E=\sum_{j=1}^{N}a_jx_j \tag{2.49}$$

在约束条件

$$\sum_{r=1}^{N}b_{j,r}x_r\leqslant c_j\qquad(j=1,2,\cdots,p) \tag{2.50-a}$$

$$\sum_{r=1}^{N} b_{i,r}x_r = c_i \qquad (i=p+1,\cdots,p+q) \tag{2.50-b}$$

$$\sum_{r=1}^{N} b_{k,r}x_r \geqslant c_k \qquad (k=p+q+1,\cdots,m) \tag{2.50-c}$$

$$x_l \geqslant 0 \qquad (l=1,2,\cdots,N) \tag{2.50-d}$$

的约束下的求极值问题。其中有 q 个等式约束条件，其余为不等式约束条件，而 $c_j \geqslant 0 \quad (j=1,2,\cdots,m)$，$m$ 为约束条件的个数。

对(2.50-a)式这种约束条件，可引入松弛变量 $x_{N+j} \geqslant 0$，把它变成等式：

$$\sum_{r=1}^{N} b_{j_1 r}x_r + x_{N+j} = c_j \qquad (j=1,2,\cdots,p)$$

而对形如(2.50-c)式所示的约束条件，可引入松弛变量 $x_{N+k} \geqslant 0$，把它变成等式：

$$\sum_{r=1}^{N} b_{k_1 r}x_r - x_{N+k} = c_k \qquad (k=p+q+1,\cdots,m)$$

引进松弛变量的目的是把不等式的约束条件化为等式约束条件，以构成统一的形式，即

$$\boldsymbol{BX}=\boldsymbol{C} \tag{2.51}$$

式中：

$$\boldsymbol{X}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \\ \vdots \\ x_{N+m} \end{bmatrix}; \qquad \boldsymbol{C}=\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \\ c_{p+1} \\ \vdots \\ c_{q+p} \\ c_{q+p+1} \\ \vdots \\ c_m \end{bmatrix} \tag{2.52}$$

$B=$

$$\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,N} & 1 & \cdots & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & & & \vdots \\ b_{p,1} & b_{p,2} & \cdots & b_{p,N} & 0 & \cdots & 1 & \cdots & \cdots & 0 \\ b_{p+1,1} & b_{p+1,2} & \cdots & b_{p+1,N} & 0 & \cdots & 0 & \cdots 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & & & \vdots \\ b_{p+q,1} & b_{p+q,2} & \cdots & b_{p+q,N} & 0 & \cdots & 0 & \cdots 0 & \cdots & 0 \\ b_{p+q+1,1} & b_{p+q+1,2} & \cdots & b_{p+q+1,N} & 0 & \cdots & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,N} & 0 & \cdots & 0 & \cdots 0 & 0\cdots & -1 \end{bmatrix}$$

$$\underbrace{\qquad}_{p} \quad \underbrace{\qquad}_{q} \quad \underbrace{\qquad}_{m-p-q} \tag{2.53}$$

松弛变量的引入，并不影响目标函数 E 在约束条件限制之下求最优解，却使问题大大简化。这时线性规划问题就变为求目标函数(2.49)式，即

$$E=\sum_{j=1}^{N} a_j x_j$$

在约束条件(2.51)式

$$\boldsymbol{BX}=\boldsymbol{C}$$

$$x_l \geqslant 0 \qquad (l=1,2,\cdots,N+m)$$

限制之下的极值问题。

我们称满足(2.51)式的解叫基本解；满足(2.51)式和(2.50-d)式的解叫基本可行解；同时满足(2.51)式、(2.50-d)式又使 E 为极小的解叫最佳基本可行解。线性规划法就是从所有基本解中找出基本可行解，然后从基本可行解中找出最佳基本可行解，作为线性规划的解。单纯形法(The simplex method)是解线性规划问题的行之有效的方法。它的解法可简要地分成两步：第一步找基本可行解；第二步找最佳基本可行解。

如果限制条件(2.51)式和(2.50-d)式不矛盾的话，不难理解，这些方程在解空间中定义一个凸的多面体(或凸集)，解必定位于其中。在目标函数为极值(极小或极大)的情况下，满足目标函数的解，必然是凸多面体的一个顶点。可见，单纯形法的第一步是确定多面体之顶点，第二步就是判断哪一个顶点是满足目标函数的最佳基本可行解。

实践证明，在许多情况下，线性规划的解是唯一的。但是，如果诸约束条件是矛盾的话，线性规划问题就无解；如果最佳基本可行解让位于多面体的一个面(或一条边)，或者说 E 函数减少(或增加)的方向上多面体并不封闭，则解有无限多个。

下面我们来讨论如何应用线性规划解 L_1 范数问题。设地球物理反问题的数据方程为

$$\boldsymbol{Gm}=\boldsymbol{d}$$

若把模型参数 $m_i(i=1,2,\cdots,N)$ 视为遵从指数分布的随机变量，则欲使模型参数的概率为最大，必须有极小目标函数

$$E=\sum_{i=1}^{N}|m_i-\langle m\rangle|$$

若模型参数的均值 $\langle m\rangle=0$，则应有

$$E=\sum_{i=1}^{N}|m_i| \tag{2.54}$$

在限制条件

$$\boldsymbol{Bm}=\boldsymbol{C} \tag{2.55}$$

$$m_i\geqslant 0$$

的约束之下求极小，其中 $\boldsymbol{Bm}=\boldsymbol{C}$ 包括所有等式和引入松弛变量后，把不等式的限制条件化为等式的限制条件。

如果模型参数 $m_i(i=1,2,\cdots,N)$ 均为正，$m_i\geqslant 0$ 的条件自然满足。当 m_i 可正可负时，可设

$$m_i=x_i-y_i \tag{2.56}$$

$$x_i,y_i\geqslant 0 \qquad (i=1,2,\cdots,N)$$

这时，相应的目标函数和约束条件变为

$$\begin{aligned}&E=\sum_{i=1}^{N}(x_i+y_i)\\&\boldsymbol{BX}+\boldsymbol{BY}=\boldsymbol{C}\\&x_i,y_i\geqslant 0\qquad(i=1,2,\cdots,N)\end{aligned}\tag{2.57}$$

不言而喻，要先用线性规划解 x_i，y_i，再根据(2.56)式求 m_i。

由于将一个 m_i 变成两个为非负的未知量 x_i、y_i，自变量的数目增加了一倍，因而计算工作量也成倍地增加。

这里要再一次强调，线性规划是解欠定问题(即等式和不等式限制条件数 m 大于或等于未知数的个数 N)且未知量遵守指数分布规律时的一种好方法。在一般的LP程序中，是欠定问题还是超定问题(即 $m<N$ 还是 $m>N$)，比较容易确定，而且一旦 $m>N$，程序就自动停止工作。然而，作为随机变量的未知参数是否服从指数分布规律，只能由反演工作者自己去判断了！

线性规划的最大优点之一是可以对目标函数强加各种先验的限制条件(等式和不等式)，还包括所有模型参数或个别模型参数的限制条件。限制条件越多，无疑工作量也越大，但所求得的解也越接近于真实。

为说明线性规划在地球物理资料反演中的应用效果，我们来分析以下两个实例：

例1 大地电磁拟地震解释法

大家知道，在一维地电模型的情况下，地表观测的大地电磁阻抗

$$\begin{aligned}Z(\omega)&=\frac{E(\omega)}{H(\omega)}\\&=Z_0(\omega)\left[1+2\sum_{n=1}^{\infty}q_n\mathrm{e}^{-2nk_n^jh_n}\right]\end{aligned}$$

或

$$R(\omega)=\left[\frac{Z(\omega)}{Z_0(\omega)}-1\right]/2=\sum_{n=1}^{\infty}q_n\mathrm{e}^{-2nk_n^jh_n}\tag{2.58}$$

式中：$E(\omega)$、$H(\omega)$是频率为 ω 时在地表测得的电场和磁场分量；$Z_0(\omega)=\sqrt{-i\omega\mu_0\rho_1}$，是整个空间之电阻率为 ρ_1 的介质充满时，大地电磁之阻抗；$k_n^j=\sqrt{-i\omega_j\mu_0\sigma_n}$ 是第 n 个微层之复波数；σ_n 为该层之电导；h_n 为该层之厚度；$R(\omega)$是在地面测得的电磁波之反射函数；q_n 是反射函数 $R(\omega)$在不考虑吸收时之离散采样值；μ_0 是导磁系数，其值为 $4\pi\times10^{-7}$ H/m。

由于 q_n 是一个无限收敛的时间序列，因此在一定精度范围内，(2.58)式可以进行截断，设保留 N 项，则有

$$R(\omega_j)\approx\sum_{n=1}^{N}q_n e^{-2nk_n^j h_n}\qquad(j=1,2,\cdots,M)\tag{2.59}$$

将上式写成矩阵，得

$$\underset{(M\times1)}{\boldsymbol{R}}=\underset{(M\times N)}{\boldsymbol{P}}\ \underset{(N\times1)}{\boldsymbol{Q}}\tag{2.60}$$

式中：

$$\boldsymbol{R}=\begin{bmatrix}R_1\\R_2\\\vdots\\R_M\end{bmatrix};\qquad\boldsymbol{Q}=\begin{bmatrix}q_1\\q_2\\\vdots\\q_N\end{bmatrix}\tag{2.61}$$

设

$$p_{jn}=e^{-2n\sqrt{-i\omega_j\mu_0\sigma_n}\cdot h_n}\tag{2.62}$$

则

$$\boldsymbol{P}=\begin{bmatrix}p_{11}&p_{12}&\cdots&p_{1N}\\p_{21}&p_{22}&\cdots&p_{2N}\\\vdots&\vdots&&\vdots\\p_{M1}&p_{M2}&\cdots&p_{MN}\end{bmatrix}\tag{2.63}$$

显然，(2.60)式是一线性方程组。把 $\boldsymbol{R}$ 作为观测数据向量，$\boldsymbol{P}$ 作为数据核，则 $\boldsymbol{Q}$ 就是待求的模型。在 $M<N$ 的欠定情况，(2.60)式可用线性规划 LP 求解。

设目标函数为

$$E=\sum_{n=1}^{N}|q_n| \tag{2.64}$$

则线性规划就是在一定的等式和不等式约束条件下，求目标函数 E 的极值问题。如何设定约束条件？在大地电磁拟地震解释法中，我们以观测值 $R_j(J=1,2,\cdots,M)$ 之观测误差 δR_j 作为设定不等式限制条件的根据，即

$$R(\omega_j)+\delta R_j\geqslant\sum_{n=1}^{N}q_n\mathrm{e}^{-2nk_n^jh_n}$$

$$R(\omega_j)-\delta R_j\leqslant\sum_{n=1}^{N}q_n\mathrm{e}^{-2nk_n^jh_n} \tag{2.65}$$

也可看成是在(2.65)式的约束之下，求目标函数(2.64)式之极小。

由于 q_n 可正可负，因此，还须设

$$q_n=x_n-y_n \qquad (n=1,2,\cdots,N) \tag{2.66}$$

且

$$x_n, y_n\geqslant 0 \tag{2.67}$$

此时，(2.64)式和(2.65)式变为

$$E=\sum_{n=1}^{N}(x_n+y_n) \tag{2.68}$$

$$R(\omega_j)+\delta R_j\geqslant\sum_{n=1}^{N}x_np_{jn}-\sum_{n=1}^{N}y_np_{jn}$$

$$R(\omega_j)-\delta R_j\leqslant\sum_{n=1}^{N}x_np_{jn}-\sum_{n=1}^{N}y_np_{jn} \tag{2.69}$$

且满足(2.67)式。

注意，由于 $R(\omega_j)$ 及 δR_j 均是复数，有实部 $R^{\mathrm{r}}(\omega_j)$，δR_j^{r} 和虚部 $R^{\mathrm{I}}(\omega_j)$，δR_j^{I} 之分，所以(2.69)式还必须按实部和虚部分开表示。

$$R^{\mathrm{r}}(\omega_j)+\delta R_j^{\mathrm{r}}\geqslant\sum_{n=1}^{N}(x_n-y_n)p_{jn}^{\mathrm{r}}$$

$$R^{r}(\omega_j)-\delta R_j^{r}\leqslant\sum_{n=1}^{N}(x_n-y_n)p_{jn}^{r} \tag{2.70}$$

$$R^{1}(\omega_j)+\delta R_j^{1}\geqslant\sum_{n=1}^{N}(x_n-y_n)p_{jn}^{1}$$

$$R^{1}(\omega_j)-\delta R_j^{1}\leqslant\sum_{n=1}^{N}(x_n-y_n)p_{jn}^{1}$$

式中：p_{jn}^{r} 和 p_{jn}^{1} 分别表示复数 P_{jn}之实部和虚部。这样，大地电磁拟地震解释法之线性反演问题变成了在(2.70)式和(2.67)式限制之下，极小目标函数(2.68)式的一个线性规划问题。这里求出的是 $x_n, y_n(n=1,2,\cdots N)$，还必须根据(2.66)式计算 q_n。之后，以 q_n 为横坐标，以电磁波($\omega=1$)在介质中传播的时间为纵坐标，就可以画出大地电磁拟地震解释法的时间剖面。

由上面的讨论可以看出，由于复数的原因，不等式限制条件由 M 个成倍地增长为 $2M$ 个；又由于误差的原因，不等式限制条件又变为 $4M$ 个。而 q_n 的可正可负，又不得不将未知变量翻一番，即由 N 个变为 $2N$ 个。这样限制条件所构成的系数矩阵，就由 $M\times N$ 阶跃变为 $4M\times 2N$ 阶这样一个庞大的系数矩阵。如 $M=100, N=50$，则 LP 须解 400×100 这样一个大型的矩阵方程。显然，这是一般小型计算机无法实现的。

图 2.2 是两层地电断面之大地电磁时间剖面。图(b)中反射系数 $r_{12}>0$，而图(a)中 $r_{12}<0$。图 2.3 是 Q 型和 A 型两个三层地电断面之大地电磁拟地震解释法之时间剖面，其中(a)所示为 Q 型地电断面，而(b)为 A 型地电断面。

例 2 设 $m(x)=1.0-0.5\cos(2\pi x) \qquad (0\leqslant x\leqslant 1) \qquad (2.71)$

$g_i(x)=\exp(-(i-1)x)$，且

$$d_i=\int_0^1 g_i(x)m(x)\mathrm{d}x \qquad (i=1,2,\cdots,11) \tag{2.72}$$

这是一组人工合成的理论数据，我们把 d_i 和 $g_i(x)$ 视为第 i 个

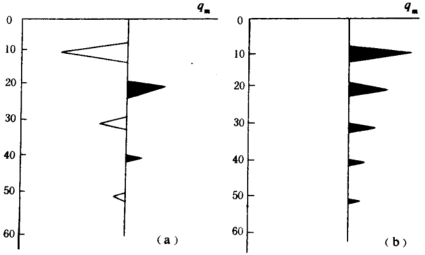


图 2.2 两层地电断面反射函数

“观测”数据和核函数，而把 $m(x)$视为模型。现在，用线性规划法，根据这 11 个观测数据计算(或反演)模型 $m(x)$。显然，这是一个连续模型的反演问题。为此，首先必须将模型参数化，即把在(0,1)区间上的 x 分为 p 个子区间，以使

$$0 \leqslant x_1 \leqslant x_2 \cdots \leqslant x_p = 1$$

这时，上述数据方程可写为

$$\sum_{i=1}^{p} G_{ji} m_i = d_j \tag{2.73}$$

式中：m_i 是第 i 个子区间中模型 $m(x)$之值，且

$$G_{ji} = \int_{x_{i-1}}^{x_i} g_j(x) \mathrm{d}x \tag{2.74}$$

是在第 i 个子区间上核函数的积分。设目标函数为：

$$E = \sum_{i=1}^{p} |m_i| \tag{2.75}$$

则线性规划的解就是在(2.73)式的约束下求 E 为极小的解。显然，

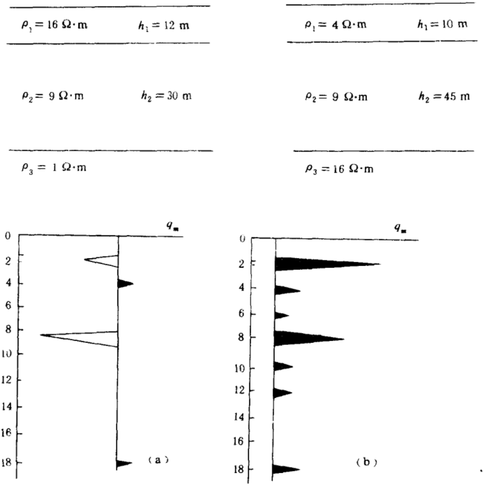


图 2.3　三层地电断面反射函数

子区间划分不同，求得的模型也不一样，如图 2.4(a)～(c)所示。

虽然，图 2.4(a)～(c)所示的离散模型可以较好地拟合 11 个观测数据，然而它与真实的连续模型相距甚远，起伏变化太大，能否设计另一种目标函数，即能减少模型沿 x 轴向的变化，又能较好地拟

合观测数据呢？下面让我们来回答这个问题。

设目标函数

$$E = \sum_{i=1}^{p} |m'_i| \tag{2.76}$$

$$\sum a_i m_i \leq \gamma$$

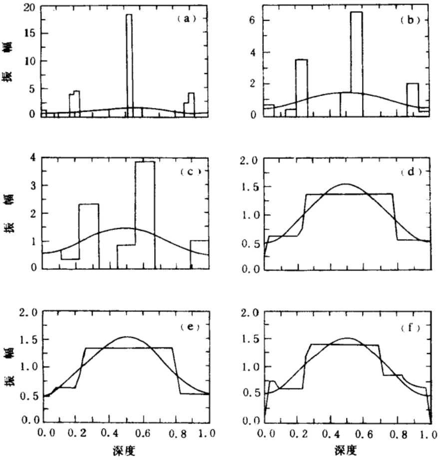


图 2.4　连续模型离散化后 LP 之反演结果

(a)～(c)最小模型，(d)～(f)最平缓模型

式中：

$m'_i=\frac{\partial m_i}{\partial x}$，是模型 m 沿 x 轴的变化率。求(2.76)式的极小，一定能确保模型沿 x 轴的起伏变化不大。图 2.4(d)～(f)是在(2.73)式的约束下，求(2.76)式的极小所得到的模型，即最平缓模型（见第四章中的有关内容）。

§7 L_∞范数解

前面讲述了地球物理资料反演中常用的两种"长度"——L_1 范数和 L_2 范数所定义的长度。除了这两种定义以外，其他范数同样可以在反演中应用。由于范数不同，自然构制出来的模型就有差异，对统计量（也许是观测数据，也许是模型参数）之统计特征的要求也不一样。这是因为，范数的定义不同，对统计量的加权值就不一样。L_∞ 突出最大者，它可以提供一种模型参数的最坏估计值。这里所谓的"最坏"是相对其他范数而言的，所以有人又称它为"极端解"(Extremal solution)或"理想解"(Idealbody solution)。

根据定义

$$L_p=\|\boldsymbol{m}\|_p=\left(\sum_{n=1}^{N}|m_i|^p\right)^{\frac{1}{p}} \tag{2.77}$$

式中：L_p 是模型 $\boldsymbol{m}$ 的 L_p 范数，所以有：

$$L_\infty=\|\boldsymbol{m}\|_\infty=\lim_{p\to\infty}\|\boldsymbol{m}\|_p=\lim_{p\to\infty}\left(\sum_{n=1}^{N}|m_i|^p\right)^{\frac{1}{p}}=\max_{1\leqslant i\leqslant N}|m_i| \tag{2.78}$$

由此可见，N 维空间中向量 $\boldsymbol{m}$ 的 L_∞范数，值等于向量 $\boldsymbol{m}$ 在 N 个坐标上投影之最大值。

设目标函数

$$\Phi=\lim_{p\to\infty}(\|\boldsymbol{m}\|_p)^p+\sum_{i=1}^{M}\lambda_i(d_i-(G_i,\boldsymbol{m})) \tag{2.79}$$

若$\frac{\partial\Phi}{\partial\boldsymbol{m}}=0$，则有

$$p(\| \boldsymbol{m} \|_p)^{p-1}=\boldsymbol{m}\sum_{i=1}^{M}\lambda_i G_i$$

不难看出，当 $\boldsymbol{m}$ 可正可负时，由于等式左端为正，所以

$$\boldsymbol{m}=\begin{cases} a & \text{当 } \Sigma\lambda_i G_i>0, \\ -a & \text{当 } \Sigma\lambda_i G_i<0 \end{cases} \tag{2.80}$$

当 $\boldsymbol{m}$ 只大于零时，是

$$\boldsymbol{m}=\begin{cases} a & \text{当 } \Sigma\lambda_i G_i>0, \\ 0 & \text{当 } \Sigma\lambda_i G_i\leqslant 0 \end{cases} \tag{2.81}$$

其结果如图 2.5 所示。

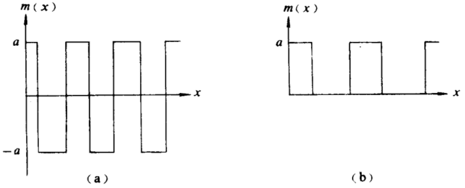


图 2.5　L_∞范数模型

(a) $\Sigma\lambda_i G_i$ 为正和负；(b) $\Sigma\lambda_i G_i$ 仅为正

显然，这里所求得的 a 是模型能拟合观测数据的最小上确界。用 L_∞范数来估计模型在定义域中的最小上确界的方法是很有用的。例如当我们用一种反演方法求得的解分辨力很低时，无法得到在某一深度(或地区)模型的可靠信息，在这种情况下，能求得模型在定义域中的最小上确界也是十分宝贵的。就拿地球密度来说吧，$\rho(r)$不可能任意大，也就是说必定存在一个最小的上“确界”。在地球中各处的

$\rho(r)$均等于或小于这个数值。而 L_∞ 范数解，就是能拟合观测数据的最小上"确界"。因为 $\rho(r)>0$，所以这时的地球密度结构成了如图2.5(b)所示的"火柴盒式"的结构。十分有趣的是，这种密度结构处处为正，任何能拟合观测数据的模型之密度都会小于或等于 a，而且，这种火柴盒式的结构是唯一的。

在后面的第四章§1中，将要讲两个数据的重力问题，其数据方程为：

$$d_1 = 1\ 833 = \int_0^1 r^2 \rho(r) \mathrm{d}r = \frac{\bar{\rho}}{3} \tag{2.82}$$

$$d_2 = 909.6 = \int_0^1 r^4 \rho(r) \mathrm{d}r = \frac{\bar{\rho} r}{2} \tag{2.83}$$

式中：$\bar{\rho}$ 为地球的平均密度；r 转动惯量，其值为 0.330 78，由(2.81)式可知，

$$\rho(r) = \begin{cases} \rho_c & \lambda_i r^2 + \lambda_i r^4 > 0 \\ 0 & \lambda_i r^2 + \lambda_i r^4 < 0 \end{cases}$$

显然，这里有三种可能：

(1) $\rho(r) = \rho_c \qquad 0 \leqslant r \leqslant 1$

(2) $\rho(r) = \begin{cases} \rho_c & 0 \leqslant r \leqslant r_c \\ 0 & r_c \leqslant r \leqslant 1 \end{cases}$

(3) $\rho(r) = \begin{cases} 0 & 0 \leqslant r \leqslant r_c \\ \rho(r) & r_c \leqslant r \leqslant 1 \end{cases}$

其结果如图2.6所示。

讨　论：

若(1)成立(见图2.6(a))，则代入数据方程(2.82)式、(2.83)式应拟合观测数据。从(2.82)式知：

$$d_1 = \frac{\bar{\rho}}{3} = \frac{\rho_c}{3}$$

即　$\bar{\rho} = \rho_c$；　而从(2.83)式知，

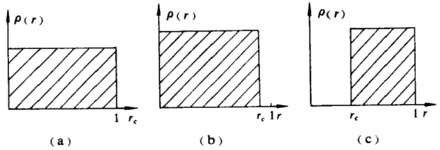


图 2.6　两个数据重力问题的三种可能模型

$$d_2=\frac{\bar{\rho}r}{2}=\frac{\rho_c}{5},\qquad 因为\quad r=0.330\ 78,$$

所以：

$$\bar{\rho}\neq\rho_c$$

显然情况(1)不成立，这不是我们需要的解。

若(2)成立(见图 2.6(b))，将其代入(2.82)式，(2.83)式得

$$d_1=\rho_c\frac{r_c^3}{3}=\frac{\bar{\rho}}{3}$$

$$d_2=\frac{r_c^5}{5}\rho_c=\frac{\bar{\rho}r}{2}$$

解之得　$r_c=\sqrt{\frac{5}{2}r}=0.909\ 4$

$$\rho_c=\frac{\bar{\rho}}{r_c^3}=\frac{5\ 500}{(0.909\ 4)^3}=7\ 313\ \mathrm{kg/m^3}$$

显然，这是一个合理的解，它说明地球的质量集中于地心这样一个重要结论。

若(3)成立(见图 2.6(c))，则

$$d_1=\frac{\overline{\rho}_c}{3}(1-r_c^3)=\frac{\overline{\rho}}{3}$$

$$d_2=\frac{\overline{\rho}_c}{3}(1-r_c^5)=\frac{\overline{\rho}r}{5}$$

以上二式显然无解，换言之，情况(3)是不可能的。

综上所述，(2)是两个数据重力问题的 L_∞ 范数解，我们知道：

$\rho_c=7\ 313\ \text{kg/m}^3$

而

$\overline{\rho}=5\ 500\ \text{kg/m}^3$

$\rho_{\max}=12\ 000\ \text{kg/m}^3$

$\rho_{地表}=2\ 800\ \text{kg/m}^3$

按照同样的方法，我们可以求出太阳和其他行星的密度问题的 L_∞ 范数解。现列表如下：

星球名称	平均密度 $\overline{\rho}(\text{kg/m}^3)$	转动惯量 r	L_∞的解 $\rho_c=\overline{\rho}/(\frac{5}{2}r)^{3/2}\ (\text{kg/m}^3)$
太　阳	1 400	0.06	24 100
火　星	4 160	0.389	4 337
木　星	1 340	0.25	2 711
土　星	680	0.22	1 667
天王星	1 555	0.23	3 566
海王星	2 230	0.29	3 612

应该指出，也可以用 L_p 解 L_∞ 范数问题，有兴趣的读者可参考有关文章。

第三章　广义反演法

前两章中，我们讨论了解线性反演问题的长度法，包括 L_1、L_2 和 L_∞ 范数解。无疑，还可以定义其他各式各样的长度，比如 L_n 范数等。但是，由于其他范数解的应用并非如此广泛，没有必要在这里加以论述了。

本章中，我们将从另一个角度，即广义逆矩阵的角度来讨论线性反演问题，并称基于广义逆矩阵建立起来的线性反演法叫广义反演法 (Generalized Inversion)，或广义线性反演法 (Generalized Linear Inversion，缩写为 GLI)。

§1　广义逆矩阵的概念

设线性反演问题

$$\underset{(M\times 1)}{\boldsymbol{d}} = \underset{(M\times N)}{\boldsymbol{G}} \underset{(N\times 1)}{\boldsymbol{m}}$$

如果把 $\boldsymbol{G}$ 看成是一个映射算子，那么正演问题就是将模型空间 (R^N) 中的模型 $\boldsymbol{m}$，通过算子 $\boldsymbol{G}$ 映射到数据空间 (R^M) 中的观测数据 $\boldsymbol{d}$ 的一种运算；而反问题则是将在数据空间中的观测数据 $\boldsymbol{d}$ 通过 $\boldsymbol{G}^{-g}$ 映射到模型空间中的模型 $\boldsymbol{m}$ 的一种运算，如图 3.1 所示。

和 $\boldsymbol{d}=\boldsymbol{Gm}$ 相应，有：

$$\boldsymbol{m}=\boldsymbol{G}^{-g}\boldsymbol{d} \tag{3.1}$$

由矩阵理论可知，若 $\boldsymbol{G}$ 是非奇异矩阵，那么 $\boldsymbol{G}^{-g}=\boldsymbol{G}^{-1}$。这里，$\boldsymbol{G}^{-1}$ 是 $\boldsymbol{G}$ 的逆矩阵，且有：

$$\boldsymbol{G}\boldsymbol{G}^{-1}=\boldsymbol{G}^{-1}\boldsymbol{G}=\boldsymbol{I} \tag{3.2}$$

式中：$\boldsymbol{I}$ 为单位矩阵；矩阵 $\boldsymbol{G}$ 的逆 $\boldsymbol{G}^{-1}$ 的求法，在一般线性代数的书中均可找到。

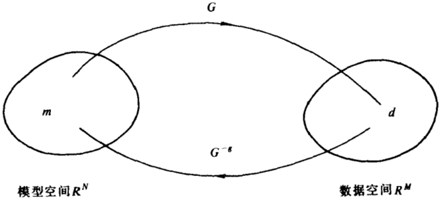


图 3.1 模型空间和数据空间的映射

在 G 是奇异矩阵的情况下，G 的逆 G^{-1} 并不存在，故我们称(3.1)式中的 G^{-g} 为矩阵 G 的广义逆。所谓广义逆，是矩阵 G 在常规意义下的逆之推广。普通逆矩阵只是广义逆矩阵的一种特殊形式。

显然，在奇异矩阵情况下，

$$GG^{-g} \neq I_M, \qquad G^{-g}G \neq I_N \tag{3.3}$$

下面，分析广义逆矩阵的形式：

(1) 当 $M=N$，且 $\det|G|\neq 0$ 时，$G^{-g}=G^{-1}$，且(3.2)式存在，反演问题的解是唯一的；

(2) 当 $M>N$，即 G 为超定系统的情况下，比较(3.1)式和(2.4)式可知

$$G^{-g}=[G^{T}G]^{-1}G^{T}; \tag{3.4}$$

且(3.3)式成立；

(3) 当 $M<N$，即 G 为欠定系统的情况下，由(3.1)式和(2.10)式知：

$$G^{-g}=G^{T}[GG^{T}]^{-1}, \tag{3.5}$$

且(3.3)式成立。

根据 Penros 的定义，凡满足以下四个条件：

(a) $GG^{-g}G=G$,

(b) $G^{-g}GG^{-g}=G^{-g}$,

(c) $(GG^{-g})^{T}=GG^{-g}$,

(d) $(G^{-g}G)^{T}=G^{-g}G$, (3.6)

的广义逆，必然是唯一的。一般把满足 Penros 四个条件的广义逆记为 G^{+}。显然，G^{-1}、最小方差解以及模型的最小欧基里德长度解(3.4)式和(3.5)式都是 Penros 的广义逆。

满足(3.6)式中四个条件的广义逆存在，当然，满足其中部分条件的广义逆也必然存在。但是，满足部分条件的广义逆并不是唯一的，因此就不是这里所说的 Penros 逆。

§2 奇异值分解(SVD)和自然逆

为了更好地了解在线性反演中应用相当普遍的奇异矩阵的奇异值分解(Singular Value Decomposition，缩写为 SVD)，让我们先从矩阵分解讲起。

(1) 若 G 为 $M\times M$ 阶实对称、非奇异矩阵，则总存在正交矩阵 U，使

$$U^{T}GU=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_M) \tag{3.7}$$

其中：λ_i 是矩阵 G 的第 i 个特征值；Λ 是由 G 的 M 个特征值组成的对角线矩阵，且

$$\Lambda=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & 0 \\ 0 & & \ddots & \\ & & & \lambda_M\end{bmatrix} \tag{3.8}$$

而$\boldsymbol{U}$是$\boldsymbol{G}$的M个特征向量组成的特征向量矩阵。显然，它是一个正交向量。

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{I}_M \tag{3.9}$$

由(3.7)式知：

$$\boldsymbol{G}=\boldsymbol{U}\Lambda\boldsymbol{U}^{\mathrm{T}} \tag{3.10}$$

这就是实对称矩阵的正交分解。任何一个实对称矩阵$\boldsymbol{G}$均可分解为三个矩阵之连乘积，第一和第三个矩阵分别为$\boldsymbol{G}$的特征向量矩阵$\boldsymbol{U}$和它的转置$\boldsymbol{U}^{\mathrm{T}}$，而第二个矩阵则是$\boldsymbol{G}$的特征值构成的对角线矩阵$\Lambda$。

(2) 如果$\boldsymbol{G}$是非奇异、非对称矩阵，那么上述正交分解不成立。可以证明，此时存在两个正交矩阵$\boldsymbol{U}$和$\boldsymbol{V}$，且

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{V}=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_M) \tag{3.11}$$

式中：Λ是$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$或$\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$特征值之正根组成的对角线矩阵，$\lambda_i>0$。而$\boldsymbol{U}$和$\boldsymbol{V}$分别为对称矩阵$\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$和$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$之特征向量组成的特征向量矩阵，且

$$\begin{aligned}&\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{I}_M\\&\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=\boldsymbol{I}_N,\end{aligned} \tag{3.12}$$

但

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{U}\neq\boldsymbol{U}^{\mathrm{T}}\boldsymbol{V}\neq\boldsymbol{I}$$

由(3.11)式知

$$\boldsymbol{G}=\boldsymbol{U}\Lambda\boldsymbol{V}^{\mathrm{T}} \tag{3.13}$$

这就是非奇异且非对称矩阵的分解。

式(3.13)和式(3.10)形式相同，但各个矩阵的含义却不尽相同。(3.13)式说明，任何一个非奇异、非对称矩阵$\boldsymbol{G}$均可分解为三个矩阵$\boldsymbol{U}$、Λ和$\boldsymbol{V}^{\mathrm{T}}$之积，其中，$\boldsymbol{U}$和$\boldsymbol{V}$分别为对称矩阵$\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$和$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$之特征向量矩阵，而$\Lambda$是$\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$或$\boldsymbol{G}^{\mathrm{T}}\mathrm{G}$特征值之正根$\lambda_i(\mathrm{i}=1,2,\cdots,M)$组成的对角线矩阵。

(3) 若$\boldsymbol{G}$是$M\times N$阶奇异矩阵，此时也可以进行分解，即：

$$\boldsymbol{G}=\boldsymbol{U}_r\Lambda_r\boldsymbol{V}_r^{\mathrm{T}} \tag{3.14}$$

我们称(3.14)式为奇异矩阵 $\boldsymbol{G}$ 之奇异值分解。其中：Λ_r 是由 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 或 $\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$ 之 r 个非零特征值之正根组成的对角线矩阵，即

$$\Lambda_r=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & 0 \\ 0 & & \ddots & \\ & & & \lambda_r\end{bmatrix} \tag{3.15}$$

式中：r 是矩阵 $\boldsymbol{G}$ 的秩。$\boldsymbol{U}_r$ 是矩阵 $\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$ 之 $M\times r$ 阶特征向量矩阵，而 $\boldsymbol{V}_r$ 是矩阵 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 之 $N\times r$ 阶特征向量矩阵，且它们都是半正交矩阵，即

$$\begin{aligned}&\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{U}_r=\boldsymbol{I}_r, \qquad \boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\neq\boldsymbol{I}_r,\\ &\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{V}_r=\boldsymbol{I}_r, \qquad \boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}}\neq\boldsymbol{I}_r\end{aligned} \tag{3.16}$$

现在，来证明(3.14)式。

组成一个$(M\times N)$阶方阵 S(Hermitiau 矩阵)，即

$$S=\begin{bmatrix}0 & \boldsymbol{G}\\ \boldsymbol{G}^{\mathrm{T}} & 0\end{bmatrix}\begin{matrix}M\\ N\end{matrix} \tag{3.17}$$

$$\qquad M \quad N$$

由矩阵代数可知，方阵 S 的特征值 λ_i 和特征向量 ω_i 之间满足

$$\boldsymbol{S}\omega_i=\lambda_i\omega_i \tag{3.18}$$

式中：ω_i 是 $M+N$ 维向量；λ_i 是与 ω_i 对应的特征值。若把 ω_i 分为 $\boldsymbol{U}_i$ (M 维向量)和 $\boldsymbol{V}_j$(N 维向量)两部分，则(3.18)式可改写为：

$$\begin{bmatrix}0 & \boldsymbol{G}\\ \boldsymbol{G}^{\mathrm{T}} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_i\\ \cdots\\ \boldsymbol{V}_i\end{bmatrix}=\lambda_i\begin{bmatrix}\boldsymbol{U}_i\\ \cdots\\ \boldsymbol{V}_i\end{bmatrix} \tag{3.19}$$

由(3.19)式可以看出，$(M\times N)$阶奇异矩阵 $\boldsymbol{G}$，可以找到两个特征向量 $\boldsymbol{U}_i$ 和 $\boldsymbol{V}_i$，使下式成立，即

$$\boldsymbol{G}\boldsymbol{V}_i=\lambda_i\boldsymbol{U}_i$$

$$G^{T}U_i=\lambda_i V_i \tag{3.20}$$

用 G^T 和 G 分别左乘(3.20)式第一式和第二式，并利用第二式和第一式进行变换，最后得：

$$\begin{aligned} G^{T}GV_i=\lambda_i G^{T}U_i=\lambda_i^2 V_i \qquad (i=1,2,\cdots,N) \\ GG^{T}U_i=\lambda_i GV_i=\lambda_i^2 U_i \qquad (i=1,2,\cdots,M) \end{aligned} \tag{3.21}$$

由于 G^TG 和 GG^T 都是实对称矩阵，其特征值 λ_i^2 均为正值，且特征向量矩阵 V 和 U 都是正交矩阵，即有：

$$U^{T}U=UU^{T}=I_M$$

$$V^{T}V=VV^{T}=I_N, \tag{3.22}$$

如果矩阵 G 的秩是 r，当 $i\leqslant r$ 时，$\lambda_i>0$；当 $i>r$ 时，$\lambda_i=0$。所以，矩阵 U 可以分解成 U_r 和 U_0，矩阵 V 可以分解成 V_r 和 V_0。其中 U_r 和 V_r 分别是 r 个非零特征值对应的 r 个特征向量构成的矩阵；U_0 和 V_0 是以零特征值对应的 $(M-r)$ 个特征向量 U_i 和 $(N-r)$ 个特征向量 V_i 所构成的矩阵。由(3.20)式的第一式知，这时

$$GV=U\Lambda \tag{3.23}$$

可以改写为：

$$\underset{(M\times N)}{G}\left[\underset{(N\times r)}{V_r}\ \underset{(N\times(N-r))}{V_0}\right]=\left[\underset{(M\times r)}{U_r}\ \underset{(M\times(M-r))}{U_0}\right]\left[\begin{array}{c:c}\Lambda_r & 0 \\ \hdashline 0^{T} & 0 \\ {\scriptstyle r} & {\scriptstyle (M-r)}\end{array}\right]\begin{matrix} r \\ (M-r)\end{matrix} \tag{3.24}$$

上式两端右乘以 $V^T=\begin{bmatrix}V_r^{T}\\V_0^{T}\end{bmatrix}$，并利用特征向量矩阵 V 之正交性，即 $VV^T=I$，则有：

$$GVV^{T}=[U_r\ U_0]\begin{bmatrix}\Lambda_r & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_r^{T}\\V_0^{T}\end{bmatrix}$$

所以

$$G=U_r\Lambda_r V_r^{T} \tag{3.25}$$

同理，从(3.20)式的第二式知：

$$G^{T}U=V\Lambda \tag{3.26}$$

可以改写为

$$\underset{(N\times M)}{G^{T}}\left[\underset{(M\times r)}{U_r}\ \underset{(M\times(M-r))}{U_0}\right]=\left[\underset{(N\times r)}{V_r}\ \underset{(N\times(N-r))}{V_0}\right]\begin{bmatrix}\Lambda_r & 0\\ 0^{T} & 0\end{bmatrix} \tag{3.27}$$

上式两端右乘以 $U^{T}=\begin{bmatrix}U_r^{T}\\ U_0^{T}\end{bmatrix}$，并利用 U 之正交性，即 $UU^{T}=I$，则有：

$$G^{T}UU^{T}=[V_r\ V_0]\begin{bmatrix}\Lambda_r & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}U_r^{T}\\ U_0^{T}\end{bmatrix}$$

同样可得(3.25)式，即

$$G=U_r\Lambda_rV_r^{T}$$

(3.25)式就是奇异矩阵 G 之奇异值分解，式中：

$$\Lambda_r=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & 0\\ 0 & & \ddots & \\ & & & \lambda_r\end{bmatrix} \tag{3.28}$$

以上就是 Lanczos 的奇异值分解理论，据此，任何一个($M\times N$)阶的矩阵 G，均可分解为(2.25)式，即可分解为三个矩阵 U_r，Λ_r 和 V_r^{T} 之乘积。

取：

$$G_L=V_r\Lambda_r^{-1}U_r^{T} \tag{3.29}$$

为矩阵 G 的逆算子，它被 Lanczos 称为"自然逆"(Natural Inverse)。Jackson 又称它为 Lanczos 逆。尔后，大多数学者(如 Aki)，包括 Penros 在内都把它称为广义逆。而把基于(3.29)式建立起来的解线性反演问题的方法统称为广义反演法，因而 $Gm=d$ 的解为：

$$m = G_L d = V_r \Lambda_r^{-1} U_r^T d \tag{3.30}$$

可以证明(3.29)式定义的自然逆满足 Penros 给出的四个条件。

(a) $GG_LG = U_r\Lambda_rV_r^TV_r\Lambda_r^{-1}U_r^TU_r\Lambda_rV_r^T = U_r\Lambda_rV_r^T = G$,

(b) $G_LGG_L = V_r\Lambda_r^{-1}U_r^TU_r\Lambda_rV_r^TV_r\Lambda_r^{-1}U_r^T = V_r\Lambda_r^{-1}U_r^T = G_L$,

(c) $[GG_L]^T = [U_r\Lambda_rV_r^TV_r\Lambda_r^{-1}U_r^T]^T = U_r\Lambda_rV_r^T\Lambda_r^{-1}U_r^T = GG_L$,

(d) $[G_LG]^T = [V_r\Lambda_r^{-1}U_r^TU_r\Lambda_rV_r^T]^T = V_r\Lambda_r^{-1}U_r^TU_r\Lambda_rV_r^T = G_LG$。

因此,G_L 也是广义逆 G^+。

§3 广义反演法

在这节里,我们只涉及基于 Lanczos 自然逆而建立起来的广义反演法,而不讨论基于一般广义逆(即不全部满足 Penros 定义的四个条件的逆)的所谓广义反演法。

设线性反演问题有

$$Gm = d$$

存在,根据自然逆的定义,有:

$$m = G_L d$$

下面,我们分如下四种情况分别论述之。

(1)当 $M = N = r$ 时,U_0 和 V_0 均不存在,即 U_r 和 V_r 都是标准的正交矩阵,且

$$GG_L = [U_r\Lambda_rV_r^TV_r\Lambda_r^{-1}U_r^T] = G_LG = I_r$$

因此,

$$G_L = G^{-1}$$

则用(3.30)式求得的 m 就是数据方程 $Gm = d$ 的唯一解。

(2)当 $r = N < M$ 时,$Gm = d$ 是超定方程。V_0 不复存在,但 U_0 存在,此时 Vr 是正交矩阵,即

$$V_r^TV_r = V_rV_r^T = I_r$$

而 U_r 是半正交矩阵,即

$$U_r^TU_r = I_r, \qquad U_rU_r^T = I_r,$$

这时，

$$G_L = V_r\Lambda_r^{-1}U_r^T = V_r\Lambda_r^{-2}V_r^T V_r\Lambda_r U_r^T$$
$$= (G^T G)^{-1}G^T$$

因此，在这种情况下，广义反演法的解为：

$$m = G_L d = (G^T G)^{-1}G^T d$$

这就是第二章中讲的最小方差解，且具有唯一性。

(3)当 $r=M<N$ 时，$Gm=d$ 是欠定方程。此时，U_0 不复存在，而 V_0 存在。U_r 是正交矩阵，且

$$U_r^T U_r = U_r U_r^T = I_r$$

而 V_r 是半正交矩阵，即

$$V_r^T V_r = I_r, \qquad V_r V_r^T = I_r,$$

这时，有：

$$G_L = V_r\Lambda_r^{-1}U_r^T = V_r\Lambda_r U_r^T U_r\Lambda_r^{-2}U_r^T$$
$$= G^T(GG^T)^{-1}$$

因此，广义反演法的解为：

$$m = G_L d = G^T(GG^T)^{-1}d$$

这就是欠定问题的最小长度解，而且解是唯一的。

(4)当 $r<\min(M,N)$ 时，U_0 和 V_0 都存在。因此，可以把广义反演解看成是同时在 U 空间极小 $\|d-Gm\|$ 和在 V 空间极小 $\|m\|$ 的结果。

为了帮助大家理解奇异值分解和广义逆的意义，现在分析两个简单的例子。

例 1 $m_1+m_2=2$

显然，这是一个欠定方程。用矩阵表示，则有：

$$Gm=d$$

其中：

$$G=[1\ 1];$$

$$\boldsymbol{m}=\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}; \qquad \boldsymbol{d}=[2]$$

由初等代数可知，方程有无限多解。现在可用广义逆解这一方程。由于

$$\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}=[2]$$

所以其特征值为 2，$\boldsymbol{G}$ 的奇异值为 $\sqrt{2}$，$\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$ 的特征向量矩阵 $\boldsymbol{U}=[1]$，$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 的特征向量矩阵为：

$$\boldsymbol{V}=\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

因而

$$\begin{aligned}\boldsymbol{G} &=\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^{\mathrm{T}}=[1]\,[\sqrt{2}\,]\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &=[1\ 1]\end{aligned}$$

可见，奇异值分解是正确的。

又由广义逆的定义可知：

$$\boldsymbol{G}_{\mathrm{L}}=\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{U}_r^{\mathrm{T}}=\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \end{bmatrix}[1]=\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

因而有：

$$\boldsymbol{m}=\boldsymbol{G}_{\mathrm{L}}\boldsymbol{d}=\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}[2]=\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

由图 3.2 可知，满足 $m_1+m_2=2$ 的解中，只有 $m_1=m_2=1$ 这一个解，即广义反演法的解距原点的距离为最小，即长度为最小。

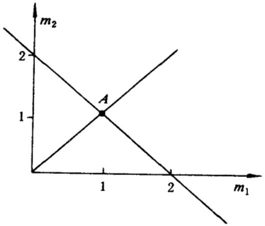


图 3.2 广义反演法解的示意图

例 2 $\begin{cases} m_1+m_2=3 \\ m_3=3 \end{cases}$

显然，这是一个欠定问题，两个方程($M=2$)，三个未知数($N=3$)。它与数据方程 $\boldsymbol{Gm}=\boldsymbol{d}$ 相应，则有：

$$\boldsymbol{G}=\begin{bmatrix}1 & 1 & 0\\0 & 0 & 1\end{bmatrix};\qquad \boldsymbol{m}=\begin{bmatrix}m_1\\m_2\\m_3\end{bmatrix};\qquad \boldsymbol{d}=\begin{bmatrix}3\\3\end{bmatrix}$$

可见，满足上述方程的解有无限多。根据 SVD 理论有：

$$\boldsymbol{GG}^{\mathrm{T}}=\begin{bmatrix}1 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 0\\0 & 1\end{bmatrix}=\begin{bmatrix}2 & 0\\0 & 1\end{bmatrix}$$

其特征值为 2 和 1；矩阵 $\boldsymbol{G}$ 的奇异值为 $\sqrt{2}$，1；对矩阵 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 来说，其奇异值也必为 $\sqrt{2}$，1。$\boldsymbol{GG}^{\mathrm{T}}$ 的特征向量为：

$$\boldsymbol{U}=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix},$$

而 G^TG 的特征向量：

$$V=\begin{bmatrix}\frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & 0\\ 0 & 1\end{bmatrix}$$

因而

$$G=U_r\Lambda_rV_r^T=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$=\begin{bmatrix}1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

则说明奇异值分解是正确的。而

$$G_L=\begin{bmatrix}\frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=\begin{bmatrix}\frac{1}{2} & 0\\ \frac{1}{2} & 1\\ 0 & 1\end{bmatrix}$$

故有：

$$m=G_Ld=\begin{bmatrix}\frac{1}{2} & 0\\ \frac{1}{2} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}3\\ 3\end{bmatrix}=\begin{bmatrix}\frac{3}{2}\\ \frac{3}{2}\\ 3\end{bmatrix}$$

可见，方程的解是 $m_1=3/2, m_2=3/2, m_3=3$。

从图 3.3 可知，在平面 $BCDF$ 上的任何点都满足 $m_1+m_2=3$；而在平面 DEF 上的任何点都满足 $m_3=3$。因此，两个平面的交线 DF 上的所有点都满足联立方程：

$$m_1+m_2=3$$
$$m_3=3$$

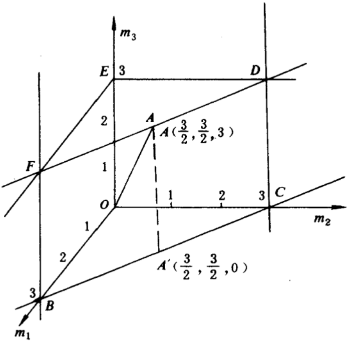


图 3.3　广义反演法示意图

但是在 DF 上距坐标原点最近的一点，即广义反演法求得的解，只能是 $m_1=\frac{3}{2}$，$m_2=\frac{3}{2}$，$m_3=3$，这就是最小长度解。

例 3 $\begin{cases} m_1+m_2=3 \\ m_1=1 \\ m_2=1 \end{cases}$

显然，这是一个超定方程，其中：

$$G=\begin{bmatrix}1 & 1\\1 & 0\\0 & 1\end{bmatrix};\qquad m=\begin{bmatrix}m_1\\m_2\end{bmatrix};\qquad d=\begin{bmatrix}3\\1\\1\end{bmatrix}$$

所以有：

$$G^{T}G=\begin{bmatrix}1&1&0\\1&0&1\end{bmatrix}\begin{bmatrix}1&1\\1&0\\0&1\end{bmatrix}=\begin{bmatrix}2&1\\1&2\end{bmatrix}$$

因而，其特征值为 3，1；G 的奇异值为 $\sqrt{3}$，1，不难求得：

$$V=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{bmatrix};\qquad U=\begin{bmatrix}\sqrt{\frac{2}{3}}&0\\\frac{1}{2}\sqrt{\frac{2}{3}}&\frac{1}{\sqrt{2}}\\\frac{1}{2}\sqrt{\frac{2}{3}}&\frac{-1}{\sqrt{2}}\end{bmatrix}$$

故有：

$$G=U_r\Lambda_rV_r^{T}=\begin{bmatrix}\sqrt{\frac{2}{3}}&0\\\frac{1}{2}\sqrt{\frac{2}{3}}&\frac{1}{\sqrt{2}}\\\frac{1}{2}\sqrt{\frac{2}{3}}&\frac{-1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{3}&0\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{bmatrix}$$

$$=\begin{bmatrix}1&1\\1&0\\0&1\end{bmatrix}$$

$$G_L=V_r\Lambda_r^{-1}U_r^{T}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{3}}&0\\0&1\end{bmatrix}^{-1}\cdot$$

$$\begin{bmatrix}\sqrt{\frac{2}{3}}&\frac{1}{2}\sqrt{\frac{2}{3}}&\frac{1}{2}\sqrt{\frac{2}{3}}\\0&\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}\frac{1}{3}&\frac{2}{3}&-\frac{1}{3}\\\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}\end{bmatrix}$$

$$m=G_L d=\begin{bmatrix}\frac{1}{3} & \frac{2}{3} & -\frac{1}{3}\\ \frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}3\\1\\1\end{bmatrix}=\begin{bmatrix}\frac{4}{3}\\ \frac{4}{3}\end{bmatrix}$$

显然，这是一个最小方差解。

§4 数据分辨矩阵

用广义反演法解线性反演问题，不但可以求得一个拟合观测数据的模型 $\boldsymbol{m}$，而且可以获得一些与观测数据 $\boldsymbol{d}$ 和模型参数 $\boldsymbol{m}$ 有关的辅助信息，例如，数据分辨矩阵(data resolution matrix)等。

假定已经求得(3.30)式所示的模型，即

$$\hat{\boldsymbol{m}}=\boldsymbol{G}_L\boldsymbol{d}=\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{d}$$

这里，用"$\hat{\boldsymbol{m}}$"表示用广义反演法构制的模型，以示和真实模型 $\boldsymbol{m}$ 之区别。试问，$\hat{\boldsymbol{m}}$ 能拟合观测数据吗？也就是说，把 $\hat{\boldsymbol{m}}$ 代入线性方程

$$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}$$

能获得与 $\boldsymbol{d}$ 相同的重建数据吗？若用 $\hat{\boldsymbol{d}}$ 表示重建数据，则有：

$$\begin{aligned}\hat{\boldsymbol{d}}&=\boldsymbol{G}\boldsymbol{G}_L\boldsymbol{d}=\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{d}=\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{d}\\&=\boldsymbol{F}\boldsymbol{d}\end{aligned}\tag{3.31}$$

式中：

$$\boldsymbol{F}=\boldsymbol{U}_r\boldsymbol{U}r^{\mathrm{T}}\tag{3.32}$$

是$(r\times r)$阶方阵，叫数据分辨矩阵(data resolution matrix)或信息密度矩阵(Information density matrix)，它是 $\hat{\boldsymbol{m}}$ 拟合观测数据 $\boldsymbol{d}$ 好坏程度的标志，如 $\boldsymbol{F}=\boldsymbol{I}_M$，则 $\hat{\boldsymbol{d}}=\boldsymbol{d}$，即 $\hat{\boldsymbol{m}}$ 完全拟合观测数据 $\boldsymbol{d}$，观测数据的方差为：

$$E=\sum_{i=1}^{M}e_i^2=[\boldsymbol{d}-\hat{\boldsymbol{d}}]^{\mathrm{T}}[\boldsymbol{d}-\hat{\boldsymbol{d}}]=0$$

如 $\boldsymbol{F}\neq\boldsymbol{I}_M$，则 $\hat{\boldsymbol{d}}\neq\boldsymbol{d}$，重建数据与观测数据之间有误差，则 $E\neq0$。

如图 3.4 所示，矩阵 $\boldsymbol{F}$ 的第 i 行中诸要素 f_{ii}越接近于 1，则 $\hat{d}_i$ 越接近于 d_i，即分辨力越高，因为

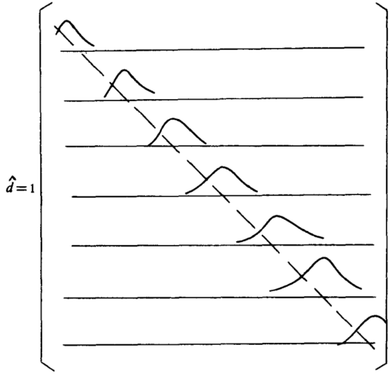


图 3.4　数据分辨矩阵

$$\hat{d}_i = \sum_{j=1}^{M} f_{i,j} d_j \tag{2.33}$$

可见，$\hat{d}_j$ 是 $d_j(j=1,2\cdots,M)$之加权平均，而权系数就是 f_{ij}，j 是矩阵 F 之元素。

如果 $f_{ij}(j=1,2\cdots,M)$的峰值虽在第 i 个位置上，但变化比较缓慢，不接近于 δ 函数，$\hat{d}_j$ 对 d_i 的分辨力不高。f_{ij}虽有峰值，但不位于第 i 个位置上，或具有多个峰值，说明数据之间存在相关，数据分辨矩阵分辨力很差，这种情况在地球物理资料反演中经常可见。

由于数据分辨矩阵 $\boldsymbol{F}$ 主对角线要素 f_{ij}表明 $\hat{d}_j$ 接近 d_i 的程度，或者 d_i 在 $\hat{d}_j$ 中所占比例(或权)的大小。因此，又定义 $\boldsymbol{F}$ 的对角线矩阵 $\boldsymbol{f}$，即

$$\boldsymbol{f}=\mathrm{diag}(\boldsymbol{F}) \tag{3.34}$$

为重要性(importance)矩阵。

由前面的讨论可知，当 $U_0=0$ 时，$\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}=\boldsymbol{I}$，或者说，当 $r=M<N$，在纯欠定的情况下，分辨矩阵 $\boldsymbol{F}$ 的分辨力最高。

显然，分辨矩阵 $\boldsymbol{F}$ 不是观测数据 $\boldsymbol{d}$ 的函数，而仅仅是数据核 $\boldsymbol{G}$ 和解反演问题时所附加先验信息的函数。在进行实际观测之前就可以把 $\boldsymbol{F}$ 计算出来，从而可以根据 $\boldsymbol{F}$ 的性态选择一组最佳的观测方式，获得一组分辨力最高的观测数据，这就是所谓的实验设计。

不难理解，当 $\boldsymbol{F}=\boldsymbol{I}$ 时，数据分辨矩阵的分辨力最高，其观测数据 $d_i(i=1,2,\cdots,M)$ 之间是独立的、无关的。此时，观测数据的利用率最高，所以

$$g_k=\sum_{j=1}^{M}\left(\sum_{i=1}^{N}U_{ki}U_{ji}-\delta_{kj}\right)^2 \tag{3.35}$$

描述了数据的利用率，g_k 越小，则利用率越高。

§5 参数分辨矩阵

和数据分辨矩阵一样，参数分辨矩阵也是广义反演法获得的另一重要辅助信息。

现在，可以设问：由广义反演法构制出来的模型 $\hat{\boldsymbol{m}}=\boldsymbol{G}_{\mathrm{L}}\boldsymbol{d}=\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{d}$ 是真正的模型 $\boldsymbol{m}$ 吗？为回答这一问题，可先将 $\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}=\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{m}$ 代入上式，则得：

$$\begin{aligned}\hat{\boldsymbol{m}}&=\boldsymbol{G}_{\mathrm{L}}\boldsymbol{G}\boldsymbol{m}=\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{m}\\&=\boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{m}\\&=\boldsymbol{R}\boldsymbol{m}\end{aligned} \tag{3.36}$$

式中：$\boldsymbol{R}$ 是 $(N\times N)$ 阶方阵，称之为参数分辨矩阵(parameter resolution matrix)或模型分辨矩阵(model resolation matrix)。它是用广义反演法构制的模型 $\hat{\boldsymbol{m}}$ 和真正地球物理模型 $\boldsymbol{m}$ 接近程度的一种重要标志。

当 $\boldsymbol{R}=\boldsymbol{I}_N$ 时，$\hat{\boldsymbol{m}}=\boldsymbol{m}$。当 $r=N<M$ 时，即在超定情况下，$\boldsymbol{V}_0=0$，才有 $\boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}}=\boldsymbol{I}_N$。这时 $\boldsymbol{R}$ 的分辨力最高。

当 $\boldsymbol{V}_0=0$ 存在时，$\boldsymbol{R}=\boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}}\neq\boldsymbol{I}_N$。所以 $\hat{\boldsymbol{m}}\neq\boldsymbol{m}$。$\hat{\boldsymbol{m}}$ 的每一个要素 m_i，均可视为 $\boldsymbol{m}$ 各要素加权的结果。这是因为：

$$\hat{m}_i=\sum_{j=1}^{N}r_{ij}m_j \tag{3.37}$$

式中：r_{ij}是矩阵 $\boldsymbol{R}$ 第 i 行，第 j 列之元素。只有 $r_{ij}=\delta_{ij}(i,j=1,2,\cdots,N)$时，$\hat{m}_i=m$，矩阵 $\boldsymbol{R}$ 的分辨力最高。

如果$(r_{ij}(j=1,2,\cdots N)$，虽有峰值，但变化比较缓慢，或者其峰值不在 $\boldsymbol{R}$ 的主对角线上，则 $\boldsymbol{R}$ 的分辨力不高。分辨力越低，说明模型参数之间越存在相关。

和数据分辨矩阵相似，参数分辨矩阵也只是数据核 $\boldsymbol{G}$ 和反演时所加先验信息的函数，而与观测数据 $\boldsymbol{d}$ 无关。因此，$\boldsymbol{R}$ 矩阵也是实验设计的重要依据。

同样，可以定义

$$h_k=\sum_{j=1}^{N}(\sum_{i=1}^{N}V_{ki}V_{ji}-\delta_{kj})^2 \tag{3.38}$$

为分辨核。h_k 越小，$\boldsymbol{R}$ 矩阵的分辨能力越高。一般取其倒数 h_k^{-1}作为分辨能力的(欲称分辨力)的定量度量。

§6 特征值和特征向量的应用

在前两节中，曾讲述了特征向量 $\boldsymbol{U}_r$ 和 $\boldsymbol{V}_r$ 在解释中的应用。作为广义反演法的辅助信息 $\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}$、$\boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}}$ 为我们揭示了观测数据之间和模型参数之间的相关程度，为更好地反演提供了一些有用的、重要的辅助信息。

在这一节中，将要论述利用广义反演法获得的另一类辅助信息——从特征值所获取的辅助信息。

1. 特征值对观测数据和模型参数的影响

将数据核矩阵 G 作奇异值分解，并代入数据方程得：

$$\hat{\boldsymbol{d}}=\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{m}$$

如用求和形式书写，则有：

$$\hat{d}_i=\sum_{j=1}^{r}U_{ij}\lambda_j\sum_{k=1}^{N}V_{jk}m_k \tag{3.39}$$

由上式可见，特征值越大，其对重建观测数据的贡献越大；相反 λ_j 越小，则对 $\hat{d}_i$ 的贡献也越小。当反演中大小特征值相差非常悬殊时，小特征值对重建观测数据几乎毫无作用，甚至将它们去掉也不会影响观测数据的重建精度。

另一方面，有

$$\hat{\boldsymbol{m}}=\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{d}$$

其求和形式为：

$$\hat{m}_i=\sum_{j=1}^{r}V_{ij}/\lambda_j\sum_{k=1}^{N}U_{jk}m_k \tag{3.40}$$

其结论和(3.39)完全相反，即特征值越小，它对构制的模型参数 $\hat{m}_i$ 影响越大。换言之，λ_j 的微小变化会导致模型参数 $\hat{m}_i$ 的巨大变化，使解变得是极不稳定。

结论必然是：注意大特征值，摒弃小特征值，是取得良好反演结果的重要途径。

2. *解的方差*

如果，观测数据具有误差 δd，当然用广义反演法所得的结果也有误差 $\boldsymbol{\delta m}$，且满足

$$\boldsymbol{\delta m}=\boldsymbol{G}_{\mathrm{L}}\boldsymbol{\delta d} \tag{3.41}$$

因此，解的协方差矩阵

$$\begin{aligned}\mathrm{cov}[\boldsymbol{m}]&=E(\boldsymbol{\delta m}\boldsymbol{\delta m}^{\mathrm{T}})\\&=E[\boldsymbol{G}_{\mathrm{L}}\boldsymbol{\delta d}\boldsymbol{\delta d}^{\mathrm{T}}\boldsymbol{G}_{\mathrm{L}}^{\mathrm{T}}]\\&=\boldsymbol{G}_{\mathrm{L}}\mathrm{cov}[\boldsymbol{d}]\boldsymbol{G}_{\mathrm{L}}^{\mathrm{T}}\end{aligned} \tag{3.42}$$

如果观测数据是统计且独立的，并有相同的方差 σ^2，则

$$\mathrm{cov}[\boldsymbol{d}]=\sigma^2\boldsymbol{I}_m$$

故

$$\mathrm{cov}[\boldsymbol{m}]=\sigma^2\boldsymbol{G}_L\boldsymbol{G}_L^T \tag{3.43}$$

单位协方差矩阵为

$$\mathrm{cov}[\boldsymbol{m}]=\boldsymbol{G}_L\boldsymbol{G}_L^T \tag{3.44}$$

讨 论：

(1)当 $M=N=r$，即 $\boldsymbol{U}_0=\boldsymbol{V}_0=0$ 时， $\boldsymbol{G}_L=\boldsymbol{G}^{-1}$，

则有：

$$\mathrm{cov}[\boldsymbol{m}]=\sigma^2\boldsymbol{G}^{-1}\boldsymbol{G}^{-1T}=\sigma^2(\boldsymbol{G}^T\boldsymbol{G})^{-1}=\sigma^2(\boldsymbol{V}\boldsymbol{\Lambda}^{-2}\boldsymbol{V}^T) \tag{3.45}$$

(2)当 $r=N\leqslant M$ 时，$\boldsymbol{V}_0=0,\boldsymbol{U}_0\neq 0,\boldsymbol{G}_L=(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T$，

则有：

$$\begin{aligned}\mathrm{cov}[\boldsymbol{m}]&=\sigma^2(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{G})^{-1}=\sigma^2(\boldsymbol{G}^T\boldsymbol{G})^{-1}\\&=\sigma^2(\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^T)\end{aligned} \tag{3.46}$$

(3)当 $r=M<N$ 时，$\boldsymbol{U}_0=0,\boldsymbol{V}_0\neq 0,\boldsymbol{G}_L=\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G})^T)^{-1}$，

则有：

$$\begin{aligned}\mathrm{cov}[\boldsymbol{m}]&=\sigma^2\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G})^T)^{-1}(\boldsymbol{G}\boldsymbol{G}^T)^{-1}\boldsymbol{G}=\sigma^2\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G})^T)^{-2}\boldsymbol{G}\\&=\sigma^2\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^T\end{aligned} \tag{3.47}$$

从(3.45)，(3.46)，(3.47)可知，它们的一般形式为：

$$\mathrm{Var}[\boldsymbol{m}]_i=\sigma^2\sum_{k=1}^{r}V_{ik}^2/\lambda_k^2 \tag{3.48}$$

显然，λ_k 越小，所对应的模型参数 $\boldsymbol{m}$ 方差越大。为了减少解的方差影响，Wiggins(1972)建议：删去一些小的特征值，因为小特征值与高分辨力相应。删去一些小特征值就意味着以牺牲一些分辨力为代价来换取较小的方差。要使得分辨力高、方差又小的解，即所谓“最优解”是不可能的，只能在分辨力和方差这一对矛盾中取折衷，求得方差合理、分辨力又不低的最佳折衷解。

解的方差 $\mathrm{Var}[m_i]$ 对了解反演结果的质量有重要意义。根据广义反演法的辅助信息，特征值 $\lambda_i(i=1,2,\cdots,r)$ 和特征向量矩阵 $\boldsymbol{V}$，可

以求得待求模型参数 m_i 之方差 $\mathrm{Var}[m_i]$。无疑，这也是一种十分重要的辅助信息。

例 1 解如下联立方程

$$\begin{cases} m_1+m_2=3 \\ \quad m_3=1 \\ -m_3=1 \end{cases}$$

显然，这是三个未知数三个方程式联立方程。其中：m_1、m_2 只与第一式有关，各有无限多解，而 m_3 与第二、三式有关，它们却是矛盾方程，无一般意义下的解，现在用广义反演法解之，并分析由此而获得的一些辅助信息，如数据分辨矩阵、参数分辨矩阵和解的方差等。

若将上式写成矩阵，即

$$\boldsymbol{Gm}=\boldsymbol{d}$$

式中：

$$\boldsymbol{G}=\begin{bmatrix}1 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & -1\end{bmatrix};\quad \boldsymbol{d}=\begin{bmatrix}3\\ 1\\ -1\end{bmatrix};\quad \boldsymbol{m}=\begin{bmatrix}m_1\\ m_2\\ m_3\end{bmatrix}$$

因此有：

$$\boldsymbol{GG}^{\mathrm{T}}=\begin{bmatrix}2 & 0 & 0\\ 0 & 1 & -1\\ 0 & -1 & 1\end{bmatrix};\quad \boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}=\begin{bmatrix}1 & 1 & 0\\ 1 & 1 & 0\\ 0 & 0 & 2\end{bmatrix}$$

由于 $\boldsymbol{GG}^{\mathrm{T}}$ 和 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 的特征值完全相同，不难求得它们分别为 2，2，0。因此，矩阵 $\boldsymbol{G}$ 之奇异值分别是：

$$\lambda_1=\sqrt{2};\quad \lambda_2=\sqrt{2};\quad \lambda_3=0$$

与对称矩阵 $\boldsymbol{GG}^{\mathrm{T}}$ 和 $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ 相对应的特征向量 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 分别是，

$$\boldsymbol{U}=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix};\quad \boldsymbol{V}=\begin{bmatrix}0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 1 & 0 & 0\end{bmatrix}$$

可见，两个矩阵中各有一个相同的特征向量。这是由于矩阵 GG^T 和 G^TG 的秩都是 2 的缘故。据此：

$$U_r=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{\sqrt{2}}\\ 0 & -\frac{1}{\sqrt{2}}\end{bmatrix};\qquad U_0=\begin{bmatrix}0\\ \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix};$$

$$V_r=\begin{bmatrix}0 & \frac{1}{\sqrt{2}}\\ 0 & \frac{1}{\sqrt{2}}\\ 1 & 0\end{bmatrix};\qquad V_0=\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\\ 0\end{bmatrix}$$

根据奇异值分解，可得：

$$G=U_r\Lambda_rV_r^T=\begin{bmatrix}0 & 1\\ \frac{1}{\sqrt{2}} & 0\\ -\frac{1}{\sqrt{2}} & 0\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & \sqrt{2}\end{bmatrix}\begin{bmatrix}0 & 0 & 1\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{bmatrix}$$

$$=\begin{bmatrix}1 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & -1\end{bmatrix}$$

$$G_L=V_r\Lambda_r^{-1}U_r^T=\begin{bmatrix}0 & \frac{1}{\sqrt{2}}\\ 0 & \frac{1}{\sqrt{2}}\\ 1 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & 0\\ 0 & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ 1 & 0 & 0\end{bmatrix}$$

$$=\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

因而有：

$$\hat{\boldsymbol{m}}=\boldsymbol{G}_L\boldsymbol{d}=\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}=\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$\hat{\boldsymbol{m}}$ 能拟合观测数据吗？将 $\hat{\boldsymbol{m}}$ 代入数据方程，可得：

$$\hat{\boldsymbol{d}}=\boldsymbol{G}\hat{\boldsymbol{m}}=\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}=\begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

显然，$\hat{\boldsymbol{d}}$ 与 $\boldsymbol{d}$ 并不完全相同，这是因为数据分辨矩阵

$$\boldsymbol{F}=\begin{bmatrix} 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}\begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \end{bmatrix}$$

$$=\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}\neq\boldsymbol{I}_m$$

由于 $\boldsymbol{F}$ 不是单位矩阵，所以由

$$\hat{d}=Fd=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{2} & -\frac{1}{2}\\ 0 & -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1\\2\\1\end{bmatrix}=\begin{bmatrix}1\\ \frac{1}{2}\\ -\frac{1}{2}\end{bmatrix}$$

表明，$\hat{d}=d_1$，$\hat{d}_2$ 和 $\hat{d}_3$ 却是 d_2 和 d_3 之加权平均，致使 $\hat{d}_2=-\hat{d}_3$，与真实的 d_2，d_3 相距甚远。

$\hat{\boldsymbol{m}}$ 与真实的模型相差多大？可将 $\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}$ 代入

$\hat{\boldsymbol{m}}=\boldsymbol{G}_L\boldsymbol{G}\boldsymbol{m}=\boldsymbol{R}\boldsymbol{m}$

中进行分析，式中：

$$\boldsymbol{R}=\boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}}=\begin{bmatrix}0 & \frac{1}{\sqrt{2}}\\ 0 & \frac{1}{\sqrt{2}}\\ 1 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 1\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{bmatrix}$$

$$=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{2} & \frac{1}{2} & 0\\ 0 & 0 & 1\end{bmatrix}\neq\boldsymbol{I}_N$$

可得：

$$\hat{\boldsymbol{m}}=\boldsymbol{R}\boldsymbol{m}=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{2} & \frac{1}{2} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}m_1\\ m_2\\ m_3\end{bmatrix}=\begin{bmatrix}\hat{m}_1\\ \hat{m}_2\\ \hat{m}_3\end{bmatrix}$$

这就说明，$\hat{m}_3=m_3$，而 $\hat{m}_1=\hat{m}_2=(m_1+m_2)/2$。由此看来，根据广义反演法，可以唯一地确定 m_3，而不能唯一地确定 m_1 和 m_2 的数值大小，只能求得它们的平均值。

至于 $\boldsymbol{m}$ 的方差：

$$\mathrm{cov}[\boldsymbol{m}]=\sigma^2\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^{\mathrm{T}}$$

$$=\sigma^2\begin{bmatrix}0 & \frac{1}{\sqrt{2}}\\ 0 & \frac{1}{\sqrt{2}}\\ 1 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{bmatrix}\begin{bmatrix}0 & 0 & 1\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{bmatrix}$$

$$=\sigma^2\begin{bmatrix}\frac{1}{4} & \frac{1}{4} & 0\\ \frac{1}{4} & \frac{1}{4} & 0\\ 0 & 0 & \frac{1}{2}\end{bmatrix}$$

从上式可知，在协方差矩阵中，两行(或列)完全相同，这是因为 $\hat{m}_1$ 和 $\hat{m}_2$ 是相关的，它们之间的相关系数为"+1"的缘故。

§7 分辨力高低和方差大小的测度

前面讨论了利用观测数据的方差和模型的长度为最小这一原则求取线性反演问题的长度解，下面将定义一种利用数据分辨矩阵 $\boldsymbol{F}$，参数分辨矩阵 $\boldsymbol{R}$ 和协方差矩阵 $\mathrm{cov}[\boldsymbol{m}]$ 计算模型参数的办法。

由于分辨矩阵($\boldsymbol{F}$,$\boldsymbol{R}$)接近单位矩阵时，说明其分辨力最高，因此一种最好办法是利用非对角线元素之大小(或其展伸情况)来描述分辨力之高低。现以英文 Spread 表示展伸系数，则有：

$$S_p(\boldsymbol{F})=\|\boldsymbol{F}-\boldsymbol{I}\|$$

$$=\sum_{i=1}^{M}\sum_{j=1}^{M}[F_{ij}-\delta_{ij}]^2 \tag{3.49}$$

$$S_p(\boldsymbol{R})=\|\boldsymbol{R}-\boldsymbol{I}\|$$

$$=\sum_{i=1}^{N}\sum_{j=1}^{N}[R_{ij}-\delta_{ij}]^2 \tag{3.50}$$

这种展伸准则有时也叫狄里西莱准则。而由(3.43)知：

$$\mathrm{cov}[\boldsymbol{m}]=\sigma^2\boldsymbol{G}_L\boldsymbol{G}_L{}^T$$

如果，我们考虑(3.49)，(3.50)和(3.43)式，并把目标函数写为

$$E=\alpha_1 S_p(\boldsymbol{F})+\alpha_2 S_p(\boldsymbol{R})+\alpha_3\mathrm{cov}[\boldsymbol{m}] \tag{3.51}$$

并极小之，可以得到模型的最佳解，式中，α_1，α_2，α_3 为相应项的加权系数。

讨　论

(1) $M>N=r$，即超定方程时，

$$E=S_p(\boldsymbol{F})=(\boldsymbol{F}-\boldsymbol{I})^T(\boldsymbol{F}-\boldsymbol{I})$$
$$=\boldsymbol{G}_L{}^T\boldsymbol{G}^T\boldsymbol{G}\boldsymbol{G}_L-\boldsymbol{G}\boldsymbol{G}_L-\boldsymbol{G}_L{}^T\boldsymbol{G}^T+\boldsymbol{I}$$

$S_p(\boldsymbol{F})$极小，必须满足

$$\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{G}_L{}^T}=\boldsymbol{G}^T\boldsymbol{G}\boldsymbol{G}_L-\boldsymbol{G}^T=0$$

故

$$\boldsymbol{G}^T\boldsymbol{G}\boldsymbol{G}_L=\boldsymbol{G}^T$$

如果我们以 $\boldsymbol{G}'_L$ 表示极小 $S_p(\boldsymbol{F})$所得的逆算子的话，则有：

$$\boldsymbol{G}'_L=(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T \tag{3.52}$$

可见，$\boldsymbol{G}'_L$ 和最小二乘法所求得的结果(2.4)或广义逆所求得的结果(3.4)完全一致。

(2) $N>M=r$，即纯欠定方程时，

$$E=S_p(\boldsymbol{R})=(\boldsymbol{R}-\boldsymbol{I})^T(\boldsymbol{R}-\boldsymbol{I})$$

将 $\boldsymbol{R}=\boldsymbol{G}_L\boldsymbol{G}$ 代入上式，并利用 $\boldsymbol{G}_L^T\boldsymbol{G}_L=(\boldsymbol{G}\boldsymbol{G}^T)^{-1}$，则

$$E=\boldsymbol{G}^T\boldsymbol{G}_L{}^T\boldsymbol{G}_L\boldsymbol{G}-\boldsymbol{G}_L\boldsymbol{G}-\boldsymbol{G}^T\boldsymbol{G}_L{}^T+\boldsymbol{I}$$
$$=\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T)^{-1}\boldsymbol{G}-\boldsymbol{G}_L\boldsymbol{G}-\boldsymbol{G}^T\boldsymbol{G}_L{}^T+\boldsymbol{I}$$

故

$$\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{G}^T}=(\boldsymbol{G}\boldsymbol{G}^T)^{-1}\boldsymbol{G}-\boldsymbol{G}_L{}^T=0$$

因而有

$$\boldsymbol{G}'_L=\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T)^{-1} \tag{3.53}$$

可见，由 $S_p(\boldsymbol{R})$ 之极小求出的逆和在第二章中纯欠定问题求得的逆(2.10)式，及第三章中的广义逆(3.5)式完全相同。

(3) 在(3.51)式，即目标函数的通式中，极小 $S_p(\boldsymbol{F})$、$S_p(\boldsymbol{R})$ 和 $\mathrm{cov}[\boldsymbol{m}]$ 显然是矛盾的，三者不可兼得。在 $\alpha_1=1,\alpha_2=\alpha_3=0$ 时，极小解就是(3.52)式；在 $\alpha_1=\alpha_3=0,\alpha_2=1$ 时，为(3.53)式；当 $\alpha_1=1,\alpha_2=0,\alpha_3=\in^2$，且 $\sigma^2=1$ 时，则有：

$$E=S_p(\boldsymbol{F})+\mathrm{cov}[\boldsymbol{m}]$$
$$=\boldsymbol{G}_{\mathrm{L}}{}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{G}_{\mathrm{L}}-\boldsymbol{G}\boldsymbol{G}_{\mathrm{L}}-\boldsymbol{G}_{\mathrm{L}}{}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}+\boldsymbol{I}+\in^2\boldsymbol{G}_{\mathrm{L}}\boldsymbol{G}_{\mathrm{L}}{}^{\mathrm{T}}$$

故

$$\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{G}_{\mathrm{L}}{}^{\mathrm{T}}}=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{G}_{\mathrm{L}}-\boldsymbol{G}^{\mathrm{T}}+\in^2\boldsymbol{G}_{\mathrm{L}}=0$$

由此可得：

$$\boldsymbol{G}'_{\mathrm{L}}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\in^2\boldsymbol{I})^{-1}\boldsymbol{G}^{\mathrm{T}} \tag{3.54}$$

它和第二章中的混定问题的逆之表达式(2.11)完全一样，此即阻尼最小二乘解。

综上所述，最小方差解也可以理解为极小数据分辨矩阵的特解；最小长度解，就是极小参数分辨矩阵的特解；而马夸特法解，则是极小数据分辨矩阵和模型方差之特定解，一言以蔽之，用广义反演法所求得的解，也就是在 L_2 范数意义下的解。

§8 最佳折衷解

在大多数地球物理反演问题中，矩阵 $\boldsymbol{G}$ 的条件数都很差，最大与最小奇异值有时相差几十个级次。我们知道，小的奇异值会引起模型参数的很大误差，却能保证模型参数的高分辨能力。分辨率和方差是一对矛盾，分辨率高必然方差大；反之，分辨率低、方差也小，二者不可兼得，只能取其折衷。或者以牺牲一些分辨率为代价换取较低的方差；或者以较大的方差为代价，获得较高的分辨率。

Wiggins 和 Jockson (1972)建议，用广义反演法求解时，设一个

最大允许方差 t，使

$$\mathrm{Var}[m_k]<t,\quad \mathrm{Var}[m_{k+1}]\geqslant t \tag{3.55}$$

即可截断或摒弃小于 λ_k 的特征值。这里 t 为“方差门坎”值。若特征值按大小顺序排列，即

$$\lambda_1>\lambda_2>\lambda\cdots>\lambda_k>\cdots>\lambda_r$$

其中仅保留 k 个大特征值，而截断 $(r-k)$ 个小特征值。显然，应按以下方法计算观测数据的有效自由度 q，即有：

$$\mathrm{Var}[m_k]=\sigma^2\sum_{i=1}^{q}\left(\frac{V_{ki}}{\lambda_i}\right)^2<t \tag{3.56}$$

其中：V_{ki} 是矩阵 $\boldsymbol{V}$ 之要素；去掉 $(r-q)$ 个奇异值相当于把矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 中最后 $(r-q)$ 个向量用零向量代替。因而，相对应的数据分辨矩阵 $\boldsymbol{F}$ 和参数分辨矩阵 $\boldsymbol{R}$ 都发生了变化，设为 $\boldsymbol{F}_k$ 和 $\boldsymbol{R}_k$，则有：

$$\boldsymbol{F}_k=\boldsymbol{V}_k\boldsymbol{V}_k^{\mathrm{T}},\qquad \boldsymbol{R}_k=\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{T}}$$

此时，相应的广义逆变为：

$$\boldsymbol{G}_{\mathrm{L}}=\boldsymbol{V}_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{U}_k^{\mathrm{T}}, \tag{3.57}$$

由于将小特征值截断的结果，使(3.38)式定义的分辨力降低了，而(3.56)式的方差却大大降低。图 3.5 是第 k 个参数分辨力和方差的示意图。分辨力随 k 的增加而提高(即(h_k 降低)，方差则随 k 值的增大而增大；反演时，令 k 值从小到大变化，计算 $h_k(k)$ 和 $\mathrm{Var}[m_k]$，以 h_k 和 $\mathrm{Var}[m_k]$为纵、横坐标，作出如图 3.6 所示的不同 k 值的折衷曲线，并由此选择最佳折衷 k 值。

折衷曲线的形态完全取决于矩阵 $\boldsymbol{G}$ 奇异值 $\lambda_k(k=1,2,\cdots,r)$的相对变化情况，图 3.6 是一种典型的形态。在实际地球物理资料反演中，还可以经常见到图 3.7 所示的折衷曲线的其他形态。曲线 a 说明，在最佳折衷点，稍微牺牲一点分辨力就可以换来方差的大幅度降低。因此，这是一种最佳折衷曲线，最佳 k 值很好确定；当奇异值变化缓慢时，会出现 b 型折衷曲线或其他类型的折衷曲线，无疑，这都不是我们所期望的。

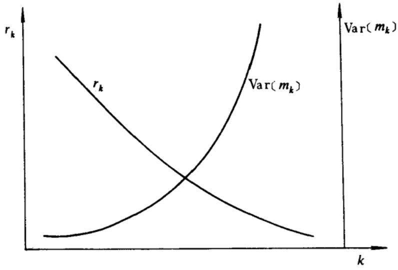


图 3.5　方差和分辨力的关系图

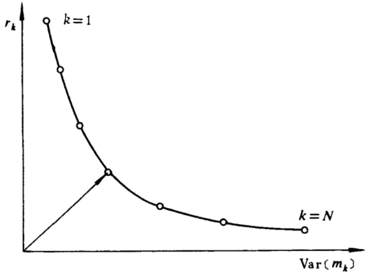


图 3.6　折衷曲线

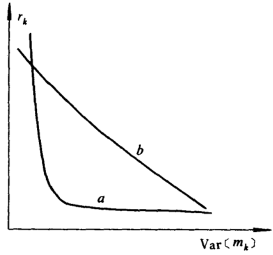


图 3.7　折衷曲线的类型

第四章 Backus-Gilbert 反演理论

在前面几章中，我们比较详细、系统地讨论了用于离散模型（含参数化的连续模型）的线性反演理论。应该说，这是一种既有较强理论性，又有广泛的适用性的有效方法技术。但是，将连续模型离散（参数）化也存在不少问题，需待进一步分析解决。

首先，实际地球物理模型十分复杂，许多情况下，并非能用有限个模型参数（或一组、几组参数向量）准确描述的。在均匀参数化（各子区间模型参数的量纲完全相同）的情况下，对模型简化的太简单，会使问题变得毫无意义，甚至根本无解；离散化太复杂，又会大大增加解的不确定性，使解释变得困难。同时，将成倍地增加计算时间，在经济上造成浪费。在非均匀参数化（指参数化模型参数具有不同的量纲）的情况下，不恰当的参数化，还可能将本来已经是线性的问题，重新非线性化；将本来简单的问题复杂化，给反演问题增加困难。

其次，连续模型参数化的结果，会降低模型参数的自由度和解的估算之置信度，从而使解的非唯一性增加和解释的难度加大。

连续介质的反演理论是反演之父——Backus 和 Gilbert 建立的，目前已形成一套完整、系统的理论（称之为 BG 理论）。在地球物理资料的反演中，已获得了广泛应用。

BG 理论包括两大部分：第一部分，在连续介质情况下，如何处理数量有限而又有误差的观测数据；第二部分，在连续介质情况下，如何处理解的非唯一性（如何从众多的非唯一的解中提取观测数据所“给予”模型的真实信息）。是盲目的，甚至“自欺欺人”把反演结果作为所谓的“最优解”接受下来，还是实事求是地对之进行合理的评价，是反演理论所必须解决的重大原则问题。这正是 BG 反演理论的

重要内容，也是 Backus 和 Gilbert 对反演理论所作出的重大贡献。

§1 在精确数据情况下连续介质的反演理论

假定地球物理模型是空间坐标 ξ 的连续函数，其数据方程可表示为：

$$d_i=\int_{r_0}^{r} g(\xi,\tau_i)m(\xi)\mathrm{d}\xi=\int_{r_0}^{r} g_i(\xi)m(\xi)\mathrm{d}\xi$$
$$=(g_i,m) \qquad (i=1,2,\cdots,M) \qquad (4.1)$$

式中：M 为观测数据的个数，M 个观测数据组成一个精确的不完整的数据集，构成 M 个积分方程；$d_i(i=1,2,\cdots,M)$是观测数据；$g(\xi,\tau_i)=g_i(\xi)=g_i$ 是核函数；$m(\xi)$是模型；τ_i 是参量$(i=1,2,\cdots,M)$；而(g_i,m)则表示内积。

下面，我们来讨论线性积分方程(4.1)的解法，即由 M 个观测数据 $d_i(i=1,2,\cdots,M)$如何求取模型 $m(\xi)$。显然，这是一个欠定问题。因为，连续模型 $m(\xi)$实际上是无限维的，欲用有限个观测数据求取需要无限个数才能准确表示的连续函数 $m(\xi)$是不可能的。按第二章中讲述的原理，必须用一些先验信息去约束或限制那些不足之处。换句话说，选择目标函数和限制条件，就是选择模型，即从千万个能拟合观测数据 $d_i(i=1,2,\cdots,M)$的模型中选择我们所需要的模型。

1. **最小模型**(smallest model)

如果我们需要一个模型参数 L_2 范数为最小的模型，则可选择

$$E=\int_{r_0}^{r}(f(\xi)m(\xi))^2\mathrm{d}\xi \qquad (4.2)$$

作为目标函数。这里，$f(\xi)$是任意选择的加权函数。在(4.1)限制下，用极小(4.2)式可求得 $m(\xi)$。这种做法并不失待求地球物理模型的基本特征，同时又能使反演问题大大简化，因而得到了广泛的应用。

在(4.1)式 M 个条件约束下，极小(4.2)式目标函数 E 的问题，就是条件极值问题。从最优化原理可知，条件极值问题必须化为求如下目标函数

$$E=\int_{r_0}^{r}[f(\xi)m(\xi)]^2\mathrm{d}\xi-\sum_{i=1}^{M}\lambda_i(d_i-\int_{r_0}^{r}g_i(\xi)m(\xi)\mathrm{d}\xi) \tag{4.3}$$

的无条件极值问题。这里的 λ_i 为拉格朗日算子。

根据变分原理中的欧拉方程，求形如

$$\int_{r_0}^{r}F(x,y(x),y'(x))\mathrm{d}x \tag{4.4}$$

为最小的积分方程的解 $y(x)$，可代之以求解微分方程，即

$$F_y-\frac{\mathrm{d}}{\mathrm{d}y}E_{y'}=0 \tag{4.5}$$

式中：

$$F_y=\frac{\partial F}{\partial y};\qquad y'=\frac{\partial y}{\partial x};\qquad F_y{}'=\frac{\partial F}{\partial y'}$$

但在(4.3)中，不存在 y'，只有 y(即 $m(\xi)$)，因此，(4.5)式就化为

$$F_y=0$$

与(4.3)式相应的(4.6)式为：

$$\frac{\partial\left\{[f(\xi)m(\xi)]^2+\sum_{i=1}^{M}\lambda_i g_i(\xi)m(\xi)\right\}}{\partial m(\xi)}=0 \tag{4.6}$$

其解为：

$$\hat{m}(\xi)=\frac{-\sum_{i=1}^{M}\lambda_i g_i(\xi)}{\alpha_i f^2(\xi)}=\frac{\sum_{i=1}^{M}\alpha_i g_i(\xi)}{f^2(\xi)} \tag{4.7}$$

式中：

$$\alpha_i=-\frac{\lambda_i}{2} \tag{4.8}$$

将(4.7)代入(4.1)，即得：

$$\begin{aligned}d_i&=\int_{r_0}^{r}g_i(\xi)\sum_{k=1}^{M}\frac{\alpha_k g_k(\xi)}{f^2(\xi)}\mathrm{d}\xi\\&=\sum_{k=1}^{M}\alpha_k\int_{r_0}^{r}\frac{g_i(\xi)g_k(\xi)}{f^2(\xi)}\mathrm{d}(\xi)=\sum_{k=1}^{M}\alpha_k G_{ik}\end{aligned} \tag{4.9}$$

式中：

$$G_{ik}=\int_{r_0}^{r}\frac{g_i(\xi)g_k(\xi)}{f^2(\xi)}\mathrm{d}\xi=\left(\frac{g_i(\xi)}{f(\xi)},\frac{g_k(\xi)}{f(\xi)}\right) \tag{4.10}$$

如果将(4.7),(4.9),(4.10)改写一下,则有

$$\hat{m}(\xi)=\frac{1}{f^2(\xi)}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{g} \tag{4.11}$$

$$\boldsymbol{d}=\boldsymbol{G\alpha} \tag{4.12}$$

式中:

$$\boldsymbol{\alpha}=\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_M\end{bmatrix};\qquad \boldsymbol{g}=\begin{bmatrix}g_1\\ g_2\\ \vdots\\ g_M\end{bmatrix} \tag{4.13}$$

$$\boldsymbol{G}=\begin{bmatrix}G_{11} & G_{12} & \cdots & G_{1M}\\ G_{21} & G_{22} & \cdots & G_{2M}\\ \vdots & \vdots & & \vdots\\ G_{M1} & G_{M2} & \cdots & G_{MM}\end{bmatrix} \tag{4.14}$$

由此不难看出,最小模型的反演步骤如下:首先,根据(4.10)计算 $G_{ik}(i,k=1,2,\cdots,M)$,并根据(4.14)组成矩阵 $\boldsymbol{G}$;其次,根据(4.12)反演求取列向量 $\boldsymbol{\alpha}$,这里 $\boldsymbol{\alpha}=\boldsymbol{G}^{-1}\boldsymbol{d}$;第三,由(4.11)式计算连续模型 $\hat{m}(\xi)$。

2. 最平缓模型(flattest model)

如果欲求一个 $m'(\xi)=\dfrac{\partial m(\xi)}{\partial(\xi)}$ 起伏最小的模型,则可选择

$$E=\int_{r_0}^{r}(f(\xi)m'(\xi))^2\mathrm{d}\xi \tag{4.15}$$

作为目标函数。

但是,数据方程(4.1)中并不包括 $m'(\xi)$。为此,对(4.1)式进行分部积分,即

$$d_i=\int_{r_0}^{r}g_i(\xi)m(\xi)\mathrm{d}\xi=m(\xi)h_i(\xi)\big|_{r_0}^{r}-\int_{r_0}^{r}m'(\xi)h_i(\xi)\mathrm{d}\xi$$

式中:

$$h_i(\xi)=\int_{r_0}^{\xi} g_i(u)\mathrm{d}u \tag{4.16}$$

$$h_i(r_0)=0$$

假设

$$P_i(\xi)=m(r)h_i(r)-d_i \tag{4.17}$$

则有：

$$\begin{aligned}P_i(\xi)&=\int_{r_0}^{r} m'(\xi)h_i(\xi)\mathrm{d}\xi\\&=(m'(\xi),h_i(\xi))\qquad (i=1,2,\cdots,M)\end{aligned} \tag{4.18}$$

显然，(4.18)式就是新的数据方程组，可以作为极小(4.15)式的约束条件。

和最小模型的求法相同，可得到 $m'(\xi)$ 的最小模型 $\hat{m}'(\xi)$，进而求得 $m(\xi)$ 的最平缓模型 $\widetilde{m}(\xi)$。其步骤如下：

首先，计算 $\dfrac{h_i(\xi)}{f(\xi)}$ 和 $\dfrac{h_k(\xi)}{f(\xi)}$ 的内积，即：

$$H_{ik}=\int_{r_0}^{r}\left(\frac{h_i(\xi)}{f(\xi)}\cdot\frac{h_k(\xi)}{f(\xi)}\right)\mathrm{d}\xi \tag{4.19}$$

并构成矩阵

$$\boldsymbol{H}=\begin{bmatrix}H_{11} & H_{12} & \cdots & H_{1M}\\ H_{21} & H_{22} & \cdots & H_{2M}\\ \vdots & \vdots & & \vdots\\ H_{M1} & H_{M2} & \cdots & H_{MM}\end{bmatrix}$$

其次，按下式

$$\boldsymbol{P}=\boldsymbol{H}\boldsymbol{\beta} \tag{4.20}$$

计算向量 $\boldsymbol{\beta}$，得 $\boldsymbol{\beta}=\boldsymbol{H}^{-1}\boldsymbol{P}$ 式中：

$$\boldsymbol{P}=\begin{bmatrix}f_1\\ f_2\\ \vdots\\ f_M\end{bmatrix};\qquad \boldsymbol{\beta}=\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_M\end{bmatrix}$$

第三，计算 $\hat{m}'(\xi)$ 的最小模型 $\hat{m}'(\xi)$，即

$$\hat{m}'(\xi)=\frac{1}{f^2(\xi)}\boldsymbol{\beta}^T\boldsymbol{h} \tag{4.21}$$

式中：

$$\boldsymbol{h}=\begin{bmatrix}h_1\\h_2\\\vdots\\h_M\end{bmatrix}$$

第四，对 $\hat{m}'(\xi)$ 积分，可得 $m(\xi)$ 的最平缓模型，即

$$\widetilde{m}(\xi)=\int_{r_0}^{\xi}\hat{m}'(u)\mathrm{d}u \tag{4.22}$$

由此可见，在求最平缓模型 $\widetilde{m}(\xi)$ 时，我们是把 $P_i(i=1,2,\cdots,M)$ 作为新的观测数据；$h_i(\xi)(i=1,2,\cdots,M)$ 为新的核函数；$m'(\xi)$ 作为新的待求最小模型。在(4.18)式的限制下，用极小目标函数(4.15)求 $\hat{m}'(\xi)$，然后对 $\hat{m}'(\xi)$ 再积分，就可求得最平缓模型 $\widetilde{m}(\xi)$。从计算过程可知，求最平缓模型必须先已知 $m(r)$，这也是一种反演过程中强加的先验信息。

不难理解，由于目标函数的差异，最平缓模型比最小模型更为平缓。在地球物理资料反演中，有时会更为有用。

3. *最光滑模型*(smoothest model)

类似地，如对(4.18)式再作一次分部积分，则得：

$$e_i=\int_{r_0}^{r}m''(\xi)k_i(\xi)\mathrm{d}\xi \tag{4.23}$$

式中：

$$k_i(\xi)=\int_{r_0}^{\xi}h_i(u)\mathrm{d}u=\int_{r_0}^{\xi}\int_{r_0}^{u}g_i(x)\mathrm{d}x\mathrm{d}u \tag{4.24}$$

$$\begin{aligned}e_i&=m'(r)k_i(r)-P_i \\ &=d_i-m(r)h_i(r)+m'(r)k_i(r)\end{aligned} \tag{4.25}$$

求最光滑模型的目标函数为：

$$E=\int_{r_0}^{r}(f(\xi)m''(\xi))^2\mathrm{d}\xi \tag{4.26}$$

其限制条件是(4.23)。按同样的方法，可得：

$$\hat{m}''(\xi)=\sum_{i=1}^{M}\frac{r_i k_i(\xi)}{f^2(\xi)} \tag{4.27}$$

进而，不难求得最光滑模型

$$\overline{m}(\xi)=\int_{r_0}^{\xi}\int_{r_0}^{u}\hat{m}''(x)\mathrm{d}x\mathrm{d}u \tag{4.28}$$

求最光滑模型的计算步骤如下：

首先，计算矩阵 $\boldsymbol{K}$ 及其要素 $K_{i\gamma}$，

$$K_{il}=\int_{r_0}^{r}\left(\frac{k_i(\xi)}{f(\xi)}\cdot\frac{k_\gamma(\xi)}{f(\xi)}\right)\mathrm{d}\xi \tag{4.29}$$

其次，按下式计算向量 $\boldsymbol{\gamma}$：

$$\boldsymbol{e}=\boldsymbol{K}\boldsymbol{\gamma} \tag{4.30}$$

且 $\boldsymbol{\gamma}=\boldsymbol{K}^{-1}\boldsymbol{e}$

式中：

$$\boldsymbol{K}=\begin{bmatrix}K_{11} & K_{12} & \cdots & K_{1M}\\ K_{21} & K_{22} & \cdots & K_{2M}\\ \vdots & \vdots & & \vdots\\ K_{M1} & K_{M2} & \cdots & K_{MM}\end{bmatrix} \tag{4.31}$$

$$\boldsymbol{e}=\begin{bmatrix}e_1\\ e_2\\ \vdots\\ e_M\end{bmatrix};\qquad \boldsymbol{\gamma}=\begin{bmatrix}\gamma_1\\ \gamma_2\\ \vdots\\ \gamma_M\end{bmatrix} \tag{4.32}$$

第三，按(4.27)式计算 $m''(\xi)$ 的最小模型 $\hat{m}''(\xi)$，即

$$\hat{m}''(\xi)=\frac{1}{f^2(\xi)}\sum_{i=1}^{M}\gamma_i k_i(\xi)=\frac{1}{f^2(\xi)}\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{k} \tag{4.33}$$

式中：

$$\boldsymbol{k}=\begin{bmatrix}k_1\\ k_2\\ \vdots\\ k_M\end{bmatrix} \tag{4.34}$$

第四，按(4.28)式计算最光滑模型 $\overline{m}(\xi)$。

由连续介质的最小模型，最平缓模型和最光滑模型的讨论中可以看出：除以观测数据方程作为限制条件外，最小模型无需另外的先验信息；而最平缓模型和最光滑模型则不同，前者需要知道在 r 处的 $m(r)$ 值，后者除 $m(r)$ 外还要知道 $m'(r)$ 的值。由于目标函数不同，限制条件各异，连续介质的这三种模型无疑不会完全一致。但是，它们都可拟合观测数据。这再一次雄辩地证明地球物理资料反演的非唯一性问题的严重性。

综上所述，三种模型有三种数据方程，即(4.1)式、(4.18)式和(4.23)式，核函数分别为 $g_i(\xi)$、$h_i(\xi)$ 和 $k_i(\xi)(i=1,2,\cdots,M)$。在反演过程中都要计算相应核函数的内积 $G_{ik},H_{ik},K_{ik}(i,k=1,2,\cdots,M)$，求(4.12)式、(4.20)式和(4.30)式之反问题，计算 α,β,γ。进而，根据各自的已知条件求出最小模型 $\hat{m}(\xi)$，最平缓模型 $\widetilde{m}(\xi)$ 和最光滑模型 $\overline{m}(\xi)$。

在解矩阵方程(4.12)式、(4.20)式和(4.30)式时，可用第二、三章讲述的任何一种方法。不过，必须认识到矩阵 $\boldsymbol{G}$、$\boldsymbol{H}$、$\boldsymbol{K}$ 都是对称正定矩阵这一重要特性，因为对称正定矩阵求逆的方法要比一般正定矩阵简单一些，在反演过程中会节省相当的计算时间。

从连续介质反演方法的步骤可以看出，不管是对最小、最平缓、还是对最光滑模型，计算对称正定矩阵的要素 G_{ij}、H_{ij}、K_{ij}是唯一的，按(4.11)式、(4.22)式和(4.28)式计算最小模型 $\hat{m}(\xi)$、最平缓模型 $\widetilde{m}(\xi)$和最光滑模型 $\overline{m}(\xi)$也不存在多解性，唯有矩阵的反演求逆才是产生非唯一性的重要原因。为缩小解的非唯一性范围，必须尽量应用好已知的先验信息。把一个线性反演问题(如这里所讲的连续介质的反演问题)化为一个简单的线性问题和一个或几个简单的非线性问题，或者化为几个简单的线性问题，也是我们解决复杂线性问题所常采用的方法。

例 1　两个数据的重力问题

为了帮助读者理解在有限个精确观测数据的情况下，连续介质模型的线性反演理论，让我们来分析一下两个数据的重力问题，其目的是用地球的质量和转动惯量求取球形对称地球的密度分布。

(1) 关于两个数据的重力问题之数据方程：

设 $\rho(r)$是球形对称地球的密度函数，r 是从地心算起的径向；m 和 c 分别为地球之质量和绕其自转轴之转动惯量。则地球的质量 m 为：

$$m=\int_0^a \rho(r)\cdot 4\pi r^2 \mathrm{d}r=\frac{4}{3}\pi a^2\bar{\rho} \tag{4.35}$$

式中：a 是地球的半径；$\bar{\rho}$ 是地球的平均密度。

设 $u=r/a$，则上式可化为

$$\frac{\bar{\rho}}{3}=\int_0^1 \rho(u)u^2 \mathrm{d}u \tag{4.36}$$

根据转动惯量的定义，对称球体绕中心轴之转动惯量为

$$\begin{aligned} c &=\int_0^a (r\sin\theta)^2\rho(r)\mathrm{d}r \\ &=\frac{8\pi}{3}\int_0^a r^4\rho(r)\mathrm{d}r \\ &=\frac{8\pi a^5}{3}\int_0^1 \rho(u)u^4\mathrm{d}u \end{aligned}$$

式中：r 是径向；θ 是 r 与地球自转轴之夹角。由普通物理学知，地球之转动惯量 c 又可表示为 $c=\varphi ma^2=\frac{4}{3}\pi a^5\bar{\rho}\cdot\varphi$，且其中 $\varphi=0.330\ 78$。所以

$$\frac{\bar{\rho}r}{2}=\int_0^1 \rho(u)u^4 \mathrm{d}u \tag{4.37}$$

若取 $\bar{\rho}=5\ 500\ \mathrm{kg/m^3}$。将有关参数代入(4.36)式，(4.37)式，则有：

$$\begin{aligned} d_1&=1\ 833=\int_0^1 \rho(u)u^2\mathrm{d}u=\bar{\rho}/3 \\ d_2&=909.5=\int_0^1 \rho(u)u^4\mathrm{d}u=\bar{\rho}r/3 \end{aligned} \tag{4.38}$$

(4.38)即是著名的两个数据重力问题的数据方程。

(2) 两个数据重力问题的最小模型：

假定(4.38)式中，观测数据 $d_1=1\ 833$，$d_2=909.5$ 是没有误差的，因此在解(4.38)式的反演问题时，是一个有限而精确数据线性积分方程的求解问题。其中：$\rho(u)$是待求的密度模型；而 u^2 和 u^4 分别是两个数据方程的核函数。

如取加权核函数 $f(u)=1$，则由 $g_1(u)=u^2$，$g_2(u)=u^4$ 知：

$$(g_1,g_2)=\frac{1}{5},\qquad (g_1,g_2)=\frac{1}{7},$$

$$(g_2,g_1)=\frac{1}{7},\qquad (g_2,g_2)=\frac{1}{9}。$$

由(4.14)知，

$$\boldsymbol{G}=\begin{bmatrix}\frac{1}{5} & \frac{1}{7}\\ \frac{1}{7} & \frac{1}{9}\end{bmatrix}$$

不难求得

$$\boldsymbol{G}^{-1}=\frac{2\ 205}{4}\begin{bmatrix}\frac{1}{9} & -\frac{1}{7}\\ -\frac{1}{7} & \frac{1}{5}\end{bmatrix},$$

故

$$\boldsymbol{\alpha}=\boldsymbol{G}^{-1}\boldsymbol{d}=\frac{2\ 205}{4}\begin{bmatrix}\frac{1}{9} & -\frac{1}{7}\\ -\frac{1}{7} & \frac{1}{5}\end{bmatrix}\begin{bmatrix}1\ 833\\ 909.5\end{bmatrix}=\begin{bmatrix}40\ 685\\ -44\ 086\end{bmatrix}$$

由此，

$$\rho(u)=\alpha_1g_1+\alpha_2g_2=40\ 685u^2-44\ 086u^4 \tag{4.39}$$

这是两个数据重力问题的最小模型。十分有趣的是，虽然这个解满足观测数据方程，能拟合观测数据，但在地心 $u=0$ 处，$\rho(u)=0$，而在地表 $u=1$ 时，$\rho(u)=-3\ 428\ \mathrm{kg/m^3}$，这在物理上却是不合理的，与已知地表和地心的地球密度知识是不一致的。显然，两个数据重力问题的最小模型不是人们可以接受的一个地球密度模型。

(3)两个数据重力问题的最平缓模型：

如前所述，为求得最平缓模型，必须附加一个先验信息，即地表的密度，设 $\rho(1)=2\ 800\ \text{kg/m}^3$。

由前面推导可知，此时的核函数为：

$$h_1(u)=\int_0^u g_1(\xi)\mathrm{d}\xi=\frac{u^3}{3},$$

$$h_2(u)=\int_0^u g_2(\xi)\mathrm{d}\xi=\frac{u^5}{5},$$

且

$$h_1(1)=\frac{1}{3};\qquad h_2(1)=\frac{1}{5}$$

观测数据方程有

$$f_1=\rho(1)h_1(1)-d_1=\frac{1}{3}(\rho(1)-\bar{\rho})$$

$$f_2=\rho(1)h_2(1)-d_2=\frac{1}{5}(\rho(1)-\frac{5}{2}\bar{\rho}r)$$

因而存在有：

$$\rho(1)-\bar{\rho}=\int_0^1\rho'(u)u^3\mathrm{d}u=2\ 700\ \text{kg/m}^3$$

$$\rho(1)-\frac{5}{2}\bar{\rho}r=\int_0^1\rho'(u)u^5\mathrm{d}u=1\ 748.2\ \text{kg/m}^3$$

这就是最平缓模型对应的数据方程，如把 $\rho'(u)$ 当作“模型”，其核函数分别为 $h_1(u)=u^3, h_2(u)=u^5$。

由此

$$\boldsymbol{H}=\begin{bmatrix}\frac{1}{7} & \frac{1}{9}\\ \frac{1}{9} & \frac{1}{11}\end{bmatrix};\qquad \boldsymbol{H}^{-1}=1\ 559.25\begin{bmatrix}\frac{1}{11} & -\frac{1}{9}\\ -\frac{1}{9} & \frac{1}{7}\end{bmatrix}$$

因而有

$$\boldsymbol{\beta}=\boldsymbol{H}^{-1}\boldsymbol{f}=-1\ 559.25\begin{bmatrix}\frac{1}{11} & -\frac{1}{9}\\ -\frac{1}{9} & \frac{1}{7}\end{bmatrix}\begin{bmatrix}2\ 700\\ 1\ 748.2\end{bmatrix}=\begin{bmatrix}-79\ 849\\ 78\ 363\end{bmatrix}$$

故

$$\rho'(u)=\beta_1 h_1+\beta_2 h_2=\beta_1 u^3+\beta_2 u^5.$$

并对 $\rho'(u)$ 积分得：

$$\rho(u)=c+\frac{\beta_1}{4}u^4+\frac{\beta_2}{6}u^6=c-19\ 962u^4+13\ 060u^6$$

利用已知先验信息 $\rho_1(1)=2\ 800\ \mathrm{kg/m^3}$，得：

$$\rho(u)=9\ 702-19\ 962u^4+13\ 060u^6 \tag{4.40}$$

(4.40)式就是两个数据重力问题的最平缓模型。由此可见，地球的密度值，此时从地心到地表缓慢减小。地心 $\rho(0)=9\ 702\ \mathrm{kg/m^3}$；地表 $\rho(1)=2\ 800\ \mathrm{kg/m^3}$。从图 4.1 可以看出，最平缓模型比最小模型好得多，它非常接近地球的真实的密度模型。

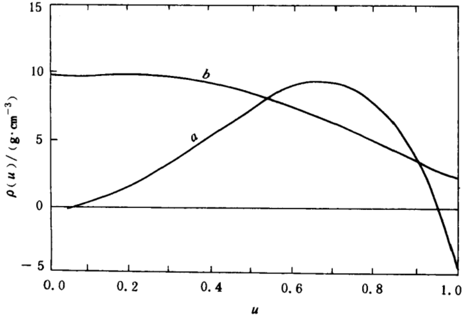


图 4.1　两个数据重力问题的最小模型和最平缓模型

按照类似的方法，我们可以求得两个数据重力问题的最光滑模型。

§2 在观测数据具有误差的情况下连续介质的反演理论

大家知道，实际观测数据都是含有误差的，只是误差大小及其所遵循的规律不同罢了。如何对待和处理有误差的观测数据，是BG反演理论的一个重点。

为了让读者更清楚地理解这里所讲述的内容，让我们先来复习一下矩阵的条件数这一概念。

1. 矩阵的条件数

设反演的数据方程

$$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m} \tag{4.41}$$

中观测数据 $\boldsymbol{d}$ 和核函数 $\boldsymbol{G}$ 的误差分别为 $\Delta\boldsymbol{d}$ 和 $\Delta\boldsymbol{G}$，试问，在这种条件下会对模型造成多大的误差 $\Delta\boldsymbol{m}$？

首先，我们讨论 $\boldsymbol{d}$ 的变化引起的 $\boldsymbol{m}$ 之变化。在线性方程时，有

$$\Delta\boldsymbol{d}=\boldsymbol{G}\Delta\boldsymbol{m} \tag{4.42}$$

对(4.41)式和(4.42)式两端取范数，则有：

$$\|\boldsymbol{d}\|\leqslant\|\boldsymbol{G}\|\cdot\|\boldsymbol{m}\|; \qquad \|\Delta\boldsymbol{m}\|\leqslant\|\boldsymbol{G}^{-1}\|\ \|\Delta\boldsymbol{d}\|$$

因而，

$$\frac{\|\Delta\boldsymbol{m}\|}{\|\boldsymbol{m}\|}\leqslant\|\boldsymbol{G}^{-1}\|\cdot\|\boldsymbol{G}\|\frac{\|\Delta\boldsymbol{d}\|}{\|\boldsymbol{d}\|}=c\frac{\|\Delta\boldsymbol{d}\|}{\|\boldsymbol{d}\|} \tag{4.43}$$

其中

$$c=\|\boldsymbol{G}^{-1}\|\cdot\|\boldsymbol{G}\| \tag{4.44}$$

称矩阵 $\boldsymbol{G}$ 的条件数。

其次，我们讨论核函数 $\boldsymbol{G}$ 的误差 $\Delta\boldsymbol{G}$ 会引起模型 $\boldsymbol{m}$ 的多大误差，由(4.41)式知，

$$(\boldsymbol{G}+\Delta\boldsymbol{G})(\boldsymbol{m}+\Delta\boldsymbol{m})=\boldsymbol{d}$$

或

$$G\Delta m=-\Delta G(m+\Delta m) \tag{4.45}$$

整理后得：

$$\Delta m=-G^{-1}\Delta G(m+\Delta m) \tag{4.46}$$

由(4.45)式、(4.46)式得

$$\frac{\|\Delta m\|}{\|m+\Delta m\|}\leqslant\|G^{-1}\|\cdot\|G\|\frac{\|\Delta G\|}{\|G\|}=c\frac{\|\Delta G\|}{\|G\|} \tag{4.47}$$

式中：c 也是条件数。

(4.43)式和(4.47)式告诉我们，模型的相对误差既与观测数据的相对误差$\frac{\|\Delta d\|}{\|d\|}$和核函数的相对误差$\frac{\|\Delta G\|}{\|G\|}$成正比，也与矩阵 G 的条件数 c 有关，即 c 越大，$\|\Delta d\|$ 或 $\|\Delta G\|$ 引起的 $\|\Delta m\|$ 越大。此时称矩阵 G 的条件很坏，会给反演带来灾难；相反 c 值越小，矩阵的条件越好，反演结果的稳定性越大。

在连续介质的反演理论中，G 是核函数的内积矩阵，内积矩阵条件数的好坏完全取决于核函数的性质及其对模型的分辨力。然而，在绝大多数地球物理资料反演中，G 的条件数都是很差的。

不难证明，当 G 为对称正定矩阵时，其条件数为

$$c=c(G)=\lambda_{\max}/\lambda_{\min} \tag{4.48}$$

式中：$\lambda_{\max}$和 $\lambda_{\min}$分别表示矩阵 G 之最大、最小特征值。经验表明，c 值有时高达 10^{30}。这就表明 d 式 G 之微小变化，也会使反演结果变得极不稳定。

2. 在观测数据具有误差的情况下连续介质的模型构制

设

$$\widetilde{d}_i=d_i^t+\delta\widetilde{d}_i \qquad (i=1,2,\cdots,M)$$

式中：d_i^t 是观测数据的真值；$\delta\widetilde{d}_i$ 是观测误差。

由于 $\delta\widetilde{d}_i$ 是一个统计量，其数值不可能准确地确定，有时连统计规律也不清楚。为理论讨论方便，而又具普遍性，我们作如下假设：

(1)每个 $\delta\widetilde{d}_i$ 均服从均值为零、方差为 σ_i^2 的高斯正态分布；

(2)观测数据的误差是不相关的，即

$$\mathrm{cov}[\delta\widetilde{d}_i,\delta\widetilde{d}_j]=0 \qquad (i\neq j)$$

如果以 σ_i 去除 $\delta\widetilde{d}_i$，即把它变成单位方差，

$$\delta d_i=\frac{\delta\widetilde{d}_i}{\sigma_i} \tag{4.49}$$

则 δd_i 也服从高斯正态分布，其均值 $E[\delta d_i]=0$，方差 $\mathrm{Var}[\delta d_i]=1$。

此时，原始数据方程

$$\widetilde{d}_i=(\widetilde{g}_i,m) \tag{4.50}$$

将变为

$$d_i=(g_i,m) \tag{4.51}$$

式中：

$$d_i=\frac{\widetilde{d}_i}{\sigma_i} \tag{4.52}$$

$$g_i=\frac{\widetilde{g}_i}{\sigma_i} \tag{4.53}$$

式(4.51)和本章"§1"中无误差观测数据方程完全相同。因此，可按相同的方法解之，即在(4.51)式 M 个条件限制之下求目标函数，即有

$$E=\int_{r_0}^{r}(f(\xi)m(\xi))^2\mathrm{d}\xi$$

之极小。就是求所谓的最小模型，和本章"§1"相同，

$$\hat{m}(\xi)=\sum_{i=1}^{M}\alpha_i\frac{g_i(\xi)}{f^2(\xi)}$$

式中：$\alpha_i(i=1,2,\cdots,M)$是如下矩阵方程的解，即

$$\boldsymbol{d}=\boldsymbol{G\alpha}$$

式中：$\boldsymbol{d},\boldsymbol{\alpha}$ 是列向量，$\boldsymbol{G}$ 是 $M\times M$ 阶方阵，其表达式和(4.13)式，(4.14)式相同，而 $f(\xi)$也是一个加权函数。

由于 $\boldsymbol{G}$ 是对称正定矩阵，可分解为

$$\boldsymbol{G}=\boldsymbol{R\Lambda R}^{\mathrm{T}} \tag{4.54}$$

式中：

$$\Lambda=\begin{bmatrix}\lambda_1 & & & 0\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_M\end{bmatrix}$$

λ_i 是 $\boldsymbol{G}$ 的第 i 个特征值，且 $\lambda_1>\lambda_2>\cdots>\lambda_M>0$，而 $\boldsymbol{R}$ 是 $\boldsymbol{G}$ 的特征向量矩阵，且满足

$$\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}=\boldsymbol{I}$$

因而

$$\boldsymbol{\alpha}=\boldsymbol{G}^{-1}\boldsymbol{d}=\boldsymbol{R}\Lambda^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{d}$$

且

$$\hat{m}(\xi)=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{g}(\xi)/f^2(\xi)$$

若取 $f(\xi)=1$，则有：

$$\begin{aligned}\hat{m}(\xi)&=\boldsymbol{d}^{\mathrm{T}}\boldsymbol{R}\Lambda^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{g}(\xi)\\&=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}=\sum_{i=1}^{M}b_i\psi_i\end{aligned}\tag{4.55}$$

式中：

$$\boldsymbol{B}=\begin{bmatrix}b_1\\b_2\\\vdots\\b_M\end{bmatrix};\qquad \boldsymbol{\Psi}=\begin{bmatrix}\psi_1\\\psi_2\\\vdots\\\psi_M\end{bmatrix}\tag{4.56}$$

且

$$b_i=\frac{1}{\sqrt{\lambda_i}}\sum_{j=1}^{M}r_{ji}d_j\tag{4.57}$$

$$\psi_i=\frac{1}{\sqrt{\lambda_i}}\sum_{j=1}^{M}r_{ji}g_j(\xi)\qquad (i=1,2,\cdots,M)\tag{4.58}$$

而 r_{ji} 是矩阵 $\boldsymbol{R}$ 的要素。

由(4.57)式和(4.58)式不难看出，ψ_i 只与核函数有关，因为 $\boldsymbol{G}$

是由核函数决定的，因而其特征向量 $r_{ji}(j=1,2,\cdots,M)$ 也必然只与核函数 $g_j(\xi)(j=1,2,\cdots,M)$ 有关。然而 b_i 则不同，它是 r_{ji} 和 d_j 之线性组合，因此，它既与核函数也与观测数据 $\boldsymbol{d}$ 有关。如果，把 b_i 看作修改后的观测数据，ψ_i 看作是坐标基，则由(4.55)式可以看出，模型 $\hat{m}(\xi)$ 就是修改后的观测数据在坐标基 $\boldsymbol{\Psi}$ 上的投影所组成的 M 维矢量。

为了说明在观测数据服从正态分布，而且互相间不相关的假设条件下，(4.55)式的意义及其在反演中的作用，让我们先讨论一下 $b_i,\psi_i(\xi)(i=1,2,\cdots,M)$ 的性质：

(1) $\psi_i(\xi)$ 是一组正交函数系

因为：

$$\int_0^\infty \psi_i(\xi)\psi_j(\xi)\mathrm{d}\xi=\int_0^\infty\left(\lambda_i^{-\frac{1}{2}}\sum_{k=1}^{M}r_{ki}g_k(\xi)\cdot\lambda_j^{-\frac{1}{2}}\sum_{l=1}^{M}r_{lj}g_l(\xi)\right)\mathrm{d}\xi$$

$$=\lambda_i^{-\frac{1}{2}}\lambda_j^{-\frac{1}{2}}\sum_{k=1}^{M}\sum_{l=1}^{M}r_{ki}\int_0^\infty g_k(\xi)g_l(\xi)\mathrm{d}\xi\cdot r_{lj}$$

$$=\lambda_i^{-\frac{1}{2}}\lambda_j^{-\frac{1}{2}}\sum_{k=1}^{M}\sum_{l=1}^{M}r_{ki}G_{kl}r_{lj}$$

$$=\frac{1}{\sqrt{\lambda_i\lambda_j}}\vec{\boldsymbol{r}}_i\cdot\boldsymbol{G}\cdot\vec{\mathbf{r}}_j$$

$$=\frac{1}{\sqrt{\lambda_i\lambda_j}}\vec{\boldsymbol{r}}_i\lambda_j\vec{\boldsymbol{r}}_j\qquad(\text{因为 }\boldsymbol{G}\boldsymbol{r}_i=\lambda_j\vec{\boldsymbol{r}}j)$$

$$=\sqrt{\frac{\lambda_j}{\lambda_i}}\vec{\boldsymbol{r}}_i\vec{\boldsymbol{r}}_i=\delta_{ij}$$

由于 $(\psi_i,\psi_j)=\delta_{ij}$，$\psi_i$ 是一组正交函数系，构成一组正交坐标基。模型 $\hat{m}(\xi)$ 就是修改后的观测数据向量 $\boldsymbol{B}$ 与这一正交坐标基的线性组合。因此，模型构制的过程，就是正交变换的过程。

(2) b_i 是统计独立的

在观测数据服从高斯分布，零平均值，方差为 σ_i^2，且互不相关的

情况下，不难证明所定义的 b_i 是统计独立的，且其标准方差为 $\lambda_i^{-\frac{1}{2}}$。

所谓 b_i 是统计独立的，就是它们之间互不相关。即：

$$\int_0^{\infty} b_i b_j \mathrm{d}\xi = \delta_{ij}$$

其证明方法和以上证明 ψ_i 是一组正交坐标基一样，这里不再赘述。下面，我们来证明 b_i 的标准方差($S \cdot d$)为 $\lambda_i^{-\frac{1}{2}}$。

$$\begin{aligned}\mathrm{Var}[b_i] &= \mathrm{Var}\left[\frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^{M} r_{ji} d_j\right] \\ &= \frac{1}{\lambda_i} \sum_{j=1}^{M} r_{ji}^2 \mathrm{var}[d_j] \\ &= \frac{1}{\lambda_i} \sum_{j=1}^{M} r_{ji}^2 \qquad (\because \quad \mathrm{Var}[d_j]=1) \\ &= \frac{1}{\lambda_i} \qquad (\because \quad \sum_{j=1}^{M} r_{ji}^2 = 1) \end{aligned} \tag{4.59}$$

故其标准方差

$$\mathrm{S} \cdot \mathrm{d}[b_i] = \frac{1}{\sqrt{\lambda_i}} \tag{4.60}$$

结合(4.57)式和(4.60)式，

$$\begin{aligned} b_i + \mathrm{S} \cdot \mathrm{d}[b_i] &= \frac{1}{\sqrt{\lambda_i}}\left[\sum_{j=1}^{M} r_{ji} d_j \pm 1\right] \\ &= \frac{1}{\sqrt{\lambda_i}}[\underline{d}_i \pm 1] \end{aligned} \tag{4.61}$$

式中：

$$\underline{d}_i = \sum_{j=1}^{M} r_{ji} d_j \tag{4.62}$$

由此可见，若 d_i 越大于 1，则相应 b_i 的精度越高；反之，若 d_i 越小于 1，b_i 的数值比其方差还小，其计算结果一定很不可靠。

(3) 观测数据拟合误差 κ^2 的计算

假定

$$\hat{m}(\xi)=\sum_{i=1}^{\bar{r}} b_i\psi_i$$

是由 $\bar{r}(\bar{r}<M)$ 个大特征值所构制出来的模型，而

$$\hat{d}_i=\int_{r_0}^{r}\hat{m}(\xi)g_i(\xi)\mathrm{d}\xi \tag{4.63}$$

是由 $\hat{m}(\xi)$ 重新计算的观测数据，因而有：

$$\kappa^2=\sum_{i=1}^{M}(d_i-\hat{d}_i)^2 \tag{4.64}$$

从统计学知，κ^2 是一个可用来判断观测数据拟合好坏的统计量。

κ^2 是一个自由度为 M 的随机变量。当 $M>5$ 时，$E(\kappa^2)\approx M$。换言之，κ^2 接近观测数据 M 时，从统计学观点讲，此时拟合最好；当 $\kappa^2 \ll M$ 时，曲线拟合过甚，$\hat{m}(\xi)$ 的细节变化来源于观测数据的误差；$\kappa^2 \gg M$ 时，曲线拟合不足，$\hat{m}(\xi)$ 中仅保留了待求模型之轮廓，而忽略了从观测数据中可以提取出来的一些细节。综上所述，选择 $\bar{r}$ 的准则是使 κ^2 接近于观测数据的个数 M。

如何计算 κ^2？根据(4.64)式，有

$$\begin{aligned}\kappa^2 &=\sum_{i=1}^{M}(d_i-\hat{d}_i)^2=\|\underline{d}-\underline{\hat{d}}\|^2\\ &=\|\boldsymbol{d}-\hat{\boldsymbol{d}}\|^2\end{aligned}$$

式中：$\underline{d}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{d}$；$\underline{\hat{d}}=\boldsymbol{R}^{\mathrm{T}}\hat{\boldsymbol{d}}$ 是旋转过后的观测数据和旋转过后的重建数据。上式之所以成立是因为：

$$\begin{aligned}\|\boldsymbol{d}-\hat{\boldsymbol{d}}\|^2 &=(\boldsymbol{d}-\hat{\boldsymbol{d}},\boldsymbol{d}-\hat{\boldsymbol{d}})\\ &=(\boldsymbol{d}-\hat{\boldsymbol{d}},\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{d}-\hat{\boldsymbol{d}}))\\ &=(\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{d}-\hat{\boldsymbol{d}}),\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{d}-\hat{\boldsymbol{d}}))\\ &=(\underline{d}-\underline{\hat{d}},\underline{d}-\underline{\hat{d}})\\ &=\|\underline{d}-\underline{\hat{d}}\|^2\end{aligned}$$

所以

$$\kappa^2=\sum_{i=1}^{M}(\underline{d}_i-\underline{\hat{d}}_i)^2 \tag{4.65}$$

由于我们仅保留了 $\bar{r}$ 个特征值，故

$$\hat{\underline{d}}_i=\underline{d}_i \qquad (i\leqslant\bar{r})$$

$$\hat{\underline{d}}_i=0 \qquad (i>\bar{r})$$

因而有

$$\kappa^2=\sum_{i=\bar{r}+1}^{M}(\underline{d}_i)^2=\sum_{i=\bar{r}+1}^{M}\lambda_i b_i^2 \tag{4.66}$$

这里，我们利用了(4.62)式，即 $\underline{d}_i=\sqrt{\lambda_i}\,b_i$。这就说明，观测数据的拟合误差 κ^2，可以用未曾应用的$(M-r)$个特征值和其对应的修改过后的观测数据 b_i 之平方乘积之和来求取。

以上我们分析了利用截断法减少小特征值在反演中所造成的不稳定性时，拟合误差 κ^2 的计算公式，下面我们再来分析用阻尼法时，κ^2 的计算方法。

加阻尼系数 ε^2 时，数据方程变为：

$$(\boldsymbol{G}+\varepsilon^2\boldsymbol{I})\boldsymbol{m}=\boldsymbol{d}$$

或

$$\boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{m}+\varepsilon^2\boldsymbol{m}=\boldsymbol{d} \tag{4.67}$$

等式两端同左乘以 $\boldsymbol{R}^{\mathrm{T}}$，则得

$$\boldsymbol{\Lambda}\underline{\boldsymbol{m}}+\varepsilon^2\underline{\boldsymbol{m}}=\underline{\boldsymbol{d}}$$

将上式(矩阵)展开，有

$$\underline{m}_i=\frac{\underline{d}_i}{\lambda_i+\varepsilon^2}=\frac{\underline{d}_i}{\lambda_i}\cdot\frac{\lambda_i}{\lambda_i+\varepsilon^2} \tag{4.68}$$

式中：

$$\underline{m}_i=\sum_{j=1}^{M}r_{ji}m_i \quad 或 \quad \underline{\boldsymbol{m}}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{m}$$

$$\underline{d}_i=\sum_{j=1}^{M}r_{ji}d_j \quad 或 \quad \underline{\boldsymbol{d}}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{d} \tag{4.69}$$

与(4.55)式相似，在采用阻尼最小二乘法时，有

$$\hat{m}(\xi)=\sum_{i=1}^{M}\frac{\underline{d}_i\sqrt{\lambda_i}}{\lambda_i+\varepsilon^2}\psi_i(\xi) \tag{4.70}$$

而

$$\kappa^2=\sum_{i=1}^{M}(d_i-\hat{d}_i)^2=\sum_{i=1}^{M}(\underline{d}_i-\underline{\hat{d}}_i)^2 \tag{4.71}$$

式中：

$$\hat{d}_i=\int_{r_0}^{r}g_i(\xi)\hat{m}(\xi)\mathrm{d}\xi=(g_i(\xi),\hat{m}(\xi))$$

$$\begin{aligned}\underline{d}_i&=\int_{r_0}^{r}\sum_{j=1}^{M}r_{ji}g_j\cdot m(\xi)\mathrm{d}\xi=(\sum_{j=1}^{M}r_{ji}g_j,m(\xi))\\&=(\sqrt{\lambda_i}\,\psi_i,m(\xi))\end{aligned} \tag{4.72}$$

$$\begin{aligned}\underline{\hat{d}}_i&=\int_{r_0}^{r}\sum_{j=1}^{M}r_{ji}g_j,\hat{m}(\xi)\mathrm{d}\xi=(\sum_{j=1}^{M}r_{ji}g_j,\hat{m}(\xi))\\&=(\sqrt{\lambda_i}\,\psi_i,\hat{m}(\xi))\end{aligned} \tag{4.73}$$

这里，我们利用了(4.58)式。

同理，利用(4.71)式、(4.72)和(4.73)式得：

$$\begin{aligned}\kappa^2&=\sum_{i=1}^{M}\lambda_i[(\psi_i,m(\xi))-(\psi_i,\hat{m}(\xi))]^2\\&=\sum_{i=1}^{M}\lambda_i\left[\frac{\underline{d}_i}{\sqrt{\lambda_i}}-\frac{\underline{d}_i\sqrt{\lambda_i}}{\lambda_i+\varepsilon^2}\right]^2\\&=\sum_{i=1}^{M}\underline{d}_i^2\left(\frac{\varepsilon^2}{\lambda_i+\varepsilon^2}\right)^2\end{aligned} \tag{4.74}$$

由此可见，κ^2 是阻尼因子 ε^2 的函数。在反演过程中，可以选择不同的 ε^2，通过迭代使 κ^2 接近 M。此时，认为拟合最佳可由(4.70)式所确定的模型就是待求的最小模型。

为便于理解，下面将举例说明连续介质中模型的构制。

例一 设 $m(x)=1.0-0.5\cos(2\pi x)$，$(0\leqslant x\leqslant 1)$ $g_i(x)=\exp(-(i-1)x)$，则有：

$$d_i=\int_0^1 g_i(x)m(x)\mathrm{d}x=\int_0^1\exp(-(i-1)x)(1.0-0.5\cos(2\pi x))\mathrm{d}x$$

$$(i=1,2,\cdots,11)$$

构成一组理论数据。已知这组理论数据和核函数 $g_i(x)=\exp(-(i-1)x)$，完成连续介质模型的构制，即求连续介质模型 $m(x)$。

下面，分两种情况说明反演结果：

(1) 有限、精确数据。图 4.2(a)是有限个精确的观测数据之最小模型；(b)～(f)是在不同约束条件下的最平缓模型。可以看出，虽然只有 11 个观测数据，但只要限制条件选择得当，最平缓模型仍然可以比较接近真实模型。

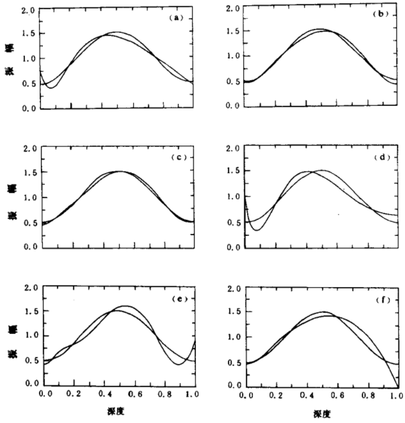


图 4.2 精确数据的反演结果

(2) 有限、非精确数据。图 4.3 是有限、非精确观测数据之反演结果。其中，第一列为最小模型，第二、三列为最平缓模型。第一行为 $\kappa^2 \ll M$；第二行 $\kappa^2 \approx M$；第三行 $\kappa^2 \gg M$。可以看出，虽然观测数据有限，M 仅为 11，而且又都有误差。但只要 κ^2 近似等于观测数据的个数 M，最平缓模型还是接近于真实模型。

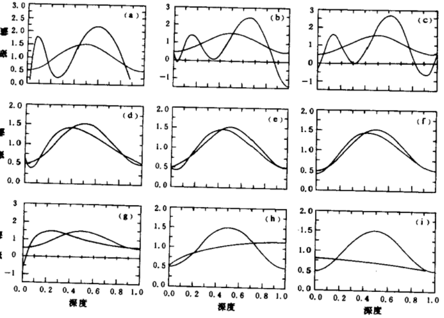


图 4.3　有误差数据的反演结果

例二　大地电磁测深曲线的反演

(1)理论曲线。图 4.4(a)～(d)是三层电阻率模型与不同的初始模型的反演结果。这里曲线 A 表示理论模型，B 表示的反演结果。由于初始模型不同，所需迭代次数和剩余误差也不相同(图中也标出了迭代次数和剩余误差)。(e)、(f)是一个八层电阻率模型及用连续介质反演的结果。

从这里可以看出，虽然是层状模型的大地电磁测深理论曲线，用连续模型反演法仍可找到一个拟合观测数据的模型。

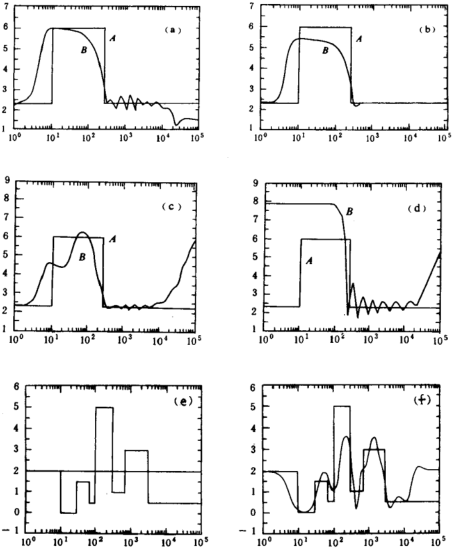


图 4.4　MT 理论数据的反演结果

(a)～(d)三层曲线；(e)、(f)八层曲线

(2)实际曲线。图 4.5 是 JDF、CAL 和 NCP 三条大地电磁测深曲线用连续介质反演方法的结果。反演时，初始模型不同，结果也不一样，但其基本形态基本一致，特别是低阻层的位置基本相同。

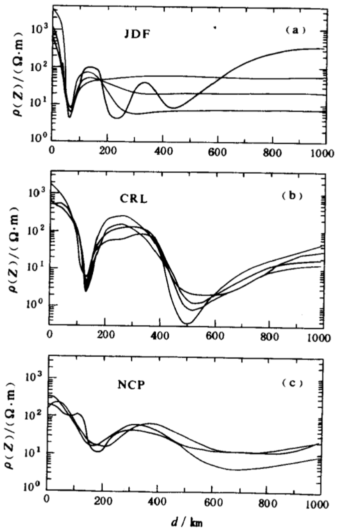


图 4.5 实测 MT 曲线的反演结果

§3 BG 线性评价(一)

既然,线性反演问题的解是非唯一的,可能有无限多个。而且,在实际工作中,不可能把所有能拟合观测数据的模型都构制出来再进行筛选。那么,我们要问,是不是用任一种反演方法构制出来的能拟合观测数据的模型就没有意义了呢?不能作出这样一个荒诞的结论。在这一节中,我们就要回答这个问题。

对待非唯一性这样一个难题,科学家们提出了两种不同的战略。一是加紧研究新的反演方法(包括线性和非线性反演方法),强加各种不同的限制条件,以缩小解的非唯一性范围,使解更逼近待求的地球物理模型;另一种是从构制出来的模型中提取所有能拟合观测数据的模型(其中也包括真实模型)的共同信息。为了提取这种信息,许多反演专家作了不懈的努力,其中最著名的、最成功的要算 Backus-Gilbert 的线性评价理论。下面,我们将详细讨论这个问题。

1. *基本理论*

观测数据方程

$$d_i=(g_i,m) \qquad (i=1,2,\cdots,M)$$

给我们一个启示:任何一个观测值d_i都可看成模型$m(\xi)$在窗口$g_i(\xi)$范围内的平均。那么,M 个观测数据 d_i 的平均可否构成某个深度 ξ_0 处的模型值 $m(\xi_0)$呢?

为回答这个问题,我们首先假设可以求得一组系数 $a_i(\xi_0)$,使它和核函数 $g_i(\xi)$之线性组合可以构成一个狄拉克 δ 函数,即

$$\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi)=\delta(\xi-\xi_0) \tag{4.75}$$

然后,再利用 $a_i(\xi_0)$与观测数据组成另一组线性组合,即

$$\begin{aligned}\sum_{i=1}^{M}a_i(\xi_0)d_i&=\sum_{i=1}^{M}a_i(\xi_0)\cdot(g_i,m)\\&=(\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi),m)\end{aligned}$$

$$=(\delta(\xi-\xi_0),m)$$

$$=m(\xi_0) \tag{4.76}$$

可见，只要在深度 ξ_0 能确定一组系数 $a_i(\xi_0)(i=1,2,\cdots,M)$，使之与 $g_i(\xi)$ 能组成一个 $\delta(\xi-\xi_0)$ 函数，即满足(4.75)，则 $a_i(\xi_0)$ 与观测数据 d_i 之线性组合，这样就可以唯一地确定深度 ξ_0 处模型的值 $m(\xi_0)$。

然而，由于M个核函数$g_i(\xi)$的线性组合是不可能确定具一个 $\delta(\xi-\xi_0)$ 函数，只能求得与 $\delta(\xi-\xi_0)$ 函数近似的函数，我们称之为平均函数 $A(\xi_0,\xi)$。

十分明显，平均函数 $A(\xi_0,\xi)$ 越接近于中心位于 ξ_0 的 δ 函数，则(4.76)就能越准确地确定 ξ_0 处的模型值 $m(\xi_0)$，即有：

$$\langle m(\xi_0)\rangle=(A(\xi_0,\xi),m(\xi)) \tag{4.77}$$

反之，$A(\xi_0,\xi)$ 越不接近于 δ 函数，中心越偏离 ξ_0，或者有多个极值存在，则 $\langle m(\xi_0)\rangle$ 就越不同于 $m(\xi_0)$。

显然，平均函数 $A(\xi,\xi_0)$ 的性态，完全取决于核函数的性质。或者说，取决于地球物理问题的本质。由于 $A(\xi,\xi_0)$ 是核函数 $g_i(\xi)(i=1,2,\cdots,M)$ 的线性组合，因此，核函数的性质不同，它随深度的衰减特性，必然会影响平均函数属性。在浅部衰减殆尽的核函数，主要决定平均函数的浅部性质；反之，衰减缓慢的核函数则影响平均函数的深部特性。就两个数据的重力问题而言，$g_1(u)=u^2$，$g_2(u)=u^4$。显然，$g_2(u)$ 的衰减比 $g_1(u)$ 快，它的分辨力比 $g_1(u)$ 高。它们线性组合的结果，$g_2(u)$ 对 $A(\xi,\xi_0)$ 的浅部特性影响大，而 $g_1(u)$ 对它的深部特性影响大。由两个核函数 u^2、u^4 线性组合，无论如何是无法获得一个 δ 函数的。因此，两个数据重力问题的平均函数的特性是很差的。

如果，我们把景物看成是“模型”，把照片看成“观测数据”的话，照相时的聚焦和光圈就是“核函数”。我们可以选择不同的聚焦和光圈以获得某一距离的景物，从而可以十分清楚地从一系列聚焦良好的照片上分清从近至远的不同景物；相反，如果照相时的聚焦和光圈

选择不当，从照片上也就不可能分辨远近不同的景物。地球物理问题和照相结果完全类似，我们就是根据在地表观测点上的一系列数据（或者同一摄影点不同聚焦的一系列照片），反演求取地球物理模型（或分辨不同景物）的。对浅层模型或近景物分辨力高，深层或远景物分辨力低，是这种问题的一个共同特性。

如把(4.77)式重新改写为：

$$\langle m(\xi_0)\rangle=\int_{r_0}^{r}A(\xi,\xi_0)m(\xi)\mathrm{d}\xi$$

则不难理解$\langle m(\xi_0)\rangle$是模型$m(\xi)$在以平均函数$A(\xi,\xi_0)$为窗口的范围内之平均值。

显然，这就是我们能从观测数据中提取出来的所有能拟合观测数据的模型（含真实模型）所包含的共同信息。之所以是共同信息，是因为所有能拟合观测数据的模型代入(4.76)以后，均可得到相同的平均值$\langle m(\xi_0)\rangle$。

设$\overline{m}(\xi)$是一个能拟合观测数据d_i的模型，故有：

$$d_i=(g_i,\overline{m})$$

将上式代入(4.76)，则有：

$$\begin{aligned}\sum_{i=1}^{M}a_i(\xi_0)d_i&=\sum_{i=1}^{M}a_i(\xi_0)\cdot(g_i,\overline{m})\\&=(\sum_{i=1}^{M}a_i(\xi_0)g_i,\overline{m})\\&=(A(\xi,\xi_0),\overline{m})\\&=\langle m(\xi_0)\rangle\end{aligned}\tag{4.78}$$

十分清楚，任何一个能够拟合观测数据的模型（当然含真实地球物理模型）上式都是成立的。因此，$\langle m(\xi_0)\rangle$是能从观测数据中提取出来的唯一信息。由此可见，由于平均函数并非δ函数，妄图从有限个观测数据d_i中提取出某一深度ξ_0的模型值$m(\xi_0)$，是不现实的，也是不可能的。能从中提取出来的唯一信息仅仅只是平均函数$A(\xi,\xi_0)$窗口范围内的模型之平均值$\langle(m(\xi_0)\rangle$。这一点读者必须十分清

楚！

如果，在从地表($\xi=0$)到足够深的点 ξ_{max} 之间取足够多的深度点 $\xi_j(j=1,2,\cdots,N)$，并对每一个深度 ξ_j 计算 $\langle m(\xi_j)\rangle$，则可以建立一个"加权平均模型"。然而，一般说来，这并不是一个可以接受的模型。这不仅仅因为，求取一个加权平均模型要耗费大量的计算时间，很不经济，而且它并不能较好地拟合观测数据。

这里还应该指出，平均函数

$$A(\xi,\xi_0)=\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi) \tag{4.79}$$

应该满足以下性质：

(1) 是规一化的，即：

$$\int_{r_0}^{r}A(\xi,\xi_0)\mathrm{d}\xi=1 \tag{4.80}$$

(2) $A(\xi,\xi_0)$的峰值应在 ξ_0 处。峰的宽度应尽可能窄，最好接近于狄拉克 δ 函数，其主叶应该大，边叶应该小，最好均为正值。

2. *加权系数 a_i 和平均函数 $A(\xi,\xi_0)$ 的确定*

(1) 第一类狄里希来准则。由于我们希望平均函数 $A(\xi,\xi_0)$尽量接近 $\delta(\xi-\xi_0)$，故此可以它们的方差为最小作为目标函数，求取待求系数 $\bar{a}_0$，这就是第一类狄里希来准则。它可以表示为以下泛函形式，即

$$E=\int_{r_0}^{r}[A(\xi,\xi_0)-\delta(\xi-\xi_0)]^2\mathrm{d}\xi \tag{4.81}$$

将(4.79)式代入上式，得：

$$E=\int_{r_0}^{r}[\sum_{i=1}^{M}a_ig_i-\delta(\xi-\xi_0)]^2\mathrm{d}\xi$$

对 a_i 求偏导数，并设其为零，则有：

$$\frac{\partial E}{\partial a_j}=2\int_{r_0}^{r}[\sum_{i=1}^{M}a_ig_i-\delta(\xi-\xi_0)]g_j\mathrm{d}\xi$$

$$=2\left\{\sum_{i=1}^{M}a_i\int_{r_0}^{r}g_ig_j\mathrm{d}\xi-\int_{r_0}^{r}\delta(\xi-\xi_0)g_j\mathrm{d}\xi\right\}=0$$

所以有：

$$\sum_{i=1}^{M} a_i \int_{r_0}^{r} g_i g_j \mathrm{d}\xi = g_j(\xi_0) \qquad (i,j=1,2,\cdots,M) \tag{4.82}$$

和以前一样，设

$$G_{ij}=(g_i,g_j) \tag{4.83}$$

则有：　　$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{A}$

式中：

$$\boldsymbol{d}=\begin{bmatrix} g_1(\xi_0) \\ g_2(\xi_0) \\ \vdots \\ g_M(\xi_0) \end{bmatrix}; \qquad \boldsymbol{A}=\begin{bmatrix} a_1(\xi_0) \\ a_2(\xi_0) \\ \vdots \\ a_M(\xi_0) \end{bmatrix} \tag{4.84}$$

$$\boldsymbol{G}=\begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1M} \\ G_{21} & G_{22} & \cdots & G_{2M} \\ \vdots & \vdots & & \vdots \\ G_{M1} & G_{M2} & \cdots & G_{MM} \end{bmatrix} \tag{4.85}$$

故

$$\boldsymbol{A}=\boldsymbol{G}^{-1}\boldsymbol{d}$$

在求得 $\boldsymbol{A}$ 之后，将其要素 $a_i(\xi_0)$ 和 $g_i(\xi)$ 线性组合，则得平均函数 $A(\xi,\xi_0)$，进而提取出 $\langle m(\xi_0)\rangle$。

为了便于理解平均函数的作用和意义，让我们分析一个简单例子。

设　$g_1(\xi)=1 \qquad 0\leqslant\xi\leqslant 1$

$$g_2(\xi)=\begin{cases} 1 & 0\leqslant\xi\leqslant\frac{1}{2} \\ 0 & \xi>\frac{1}{2} \end{cases}$$

根据(4.85)式，有：

$$G_{11}=\int_0^1 g_1(\xi)g_1(\xi)\mathrm{d}\xi=1$$

$$G_{21}=\int_0^1 g_1(\xi)g_2(\xi)\mathrm{d}\xi=\int_0^{\frac{1}{2}} g_1(\xi)g_2(\xi)\mathrm{d}\xi=G_{12}=\frac{1}{2}$$

$$G_{22}=\int_0^{\frac{1}{2}} g_2(\xi)g_2(\xi)\mathrm{d}\xi=\frac{1}{2}$$

所以有：

$$\boldsymbol{G}=\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

$$\boldsymbol{G}^{-1}=4\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{bmatrix}$$

假定我们要求 $\xi_0=3/4$ 处之平均函数，则有：

$$\boldsymbol{d}=[g_1(\xi_0)\quad g_2(\xi_0)]^{\mathrm{T}}=[1\ 0]^{\mathrm{T}}$$

所以

$$\boldsymbol{A}=\boldsymbol{G}^{-1}\boldsymbol{d}=4\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}=\begin{bmatrix}2\\ -2\end{bmatrix}=\begin{bmatrix}a_1(\xi_0)\\ a_2(\xi_0)\end{bmatrix}$$

故

$$A(\xi,\xi_0)=\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi)=\begin{cases}0 & (0\leqslant\xi\leqslant\frac{1}{2})\\ 2 & (\frac{1}{2}\leqslant\xi\leqslant 1)\end{cases}$$

显然，$A(\xi,\xi_0)$是单位模型，其面积为 1，且中心位于 3/4 处。

现在，我们再看一下 $\xi_0=1/4$ 处之 $A(\xi,\xi_0)$。

此时

$$\boldsymbol{d}=[1\ 1]^{\mathrm{T}}$$

故

$$A=G^{-1}d=4\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}0\\2\end{bmatrix}$$

因而有：

$$A(\xi,\xi_0)=\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi)=\begin{cases}2 & (0\leqslant\xi\leqslant\frac{1}{2})\\ 0 & (\frac{1}{2}\leqslant\xi\leqslant 1)\end{cases}$$

可见，$A(\xi,\xi_0)$也满足平均函数之基本条件。

从以上这个简单例子可以看出，用第一类狄里希来准则，一般总可以找到一个满足要求的平均函数 $A(\xi,\xi_0)$，遗憾的是，在这个例子中，于 $0\leqslant\xi\leqslant\frac{1}{2}$范围内（同理在$\frac{1}{2}\leqslant\xi\leqslant 1$ 范围内）的所有点上，平均函数 $A(\xi,\xi_0)$的形式完全相同，这就是说，在这一范围内，平均函数非但不接近于 δ 函数，反而成了一个门函数，致使反演结果不可能分辨出模型细节。

综上所述，第一类狄里希来准则给出的结果，仅仅是平均函数$A(\xi,\xi_0)$与$\delta(\xi,\xi_0)$之方差为最小这种意义上的"最高分辨力"。当然，准则不同，结果也不一样。

必须注意：

(a) 由于狄拉克 δ 函数之平方的积分（如 4.81）是无限大，目标函数 E 是无界的，不能直接确定。但庆幸的是，δ 函数的微分是有界的，在对E的极小化过程中，其导数是常数。因此，仍然是可以求出 $a_i(\xi_0)$，$i=1,2,\cdots,M$。

(b) 由于 E 是无界的，因此不能用来评价 $A(\xi,\xi_0)$的分辨力。但由于 $A(\xi,\xi_0)$以面积为单位，其主叶越高，宽度就越窄，边叶也必然越小，因此，可以用主叶峰值的倒数，即

$$W(\xi_0)=\frac{1}{A(\xi,\xi_0)} \tag{4.86}$$

来评价分辨力的高低。

(c) 实践证明,按第一类狄里希来准则求出的平均函数之边叶,有正有负。由于负值的出现,使其在$\langle m(\xi_0)\rangle$——即ξ_0处平均函数$A(\xi,\xi_0)$窗口范围内,模型$m(\xi)$之平均值的解释困难增大了。尽管$\langle m(\xi_0)\rangle$是唯一的信息,解释人员仍然难以获得模型$m(\xi)$在ξ_0附近的确切值。因此,从某种意义讲,这种具有负边叶的平均函数还不如图 4.6 所示的平均函数,其主叶虽较宽,但边叶无负值出现。在实际工作中,人们宁肯应用后者而不是应用具有负边叶的平均函数。

(2) 第二类狄里希来准则。为了克服用第一类狄里希来准则求取平均函数$A(\xi,\xi_0)$存在的不足,我们拟采用 Heaviside 阶跃函数,而不再是用δ函数来构筑目标函数。

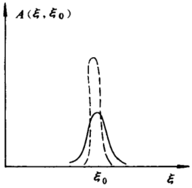


图 4.6　平均函数$A(\xi,\xi_0)$的形状

在ξ_0处,阶跃函数的定义是:

$$H(\xi-\xi_0)=\int_0^\xi \delta(u-\xi_0)\mathrm{d}u \tag{4.87}$$

大家知道,核函数的线性组合可以逼近δ函数。因此,用核函数的不定积分的线性组合,也可以逼近阶跃函数$\widetilde{H}(\xi-\xi_0)$,即

$$\widetilde{H}(\xi-\xi_0)=\sum_{i=1}^{M}a_i(\xi_0)\int_{r_0}^{\xi}g_i(u)\mathrm{d}u \tag{4.88}$$

设

$$u_i(\xi)=\int_{r_0}^{\xi}g_i(u)\mathrm{d}u \tag{4.89}$$

则

$$\widetilde{H}(\xi-\xi_0)=\sum_{i=1}^{M}a_i(\xi_0)u_i(\xi) \tag{4.90}$$

$$\delta(\xi,\xi_0)=\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi) \tag{4.91}$$

现在，我们用 $\widetilde{H}(\xi-\xi_0)$ 和 $H(\xi-\xi_0)$ 之差的平方来构筑目标函数，即

$$E=\int_{r_0}^{r}[\widetilde{H}(\xi-\xi_0)-H(\xi-\xi_0)]^2\mathrm{d}\xi$$

并极小之，将(4.90)代入上式，则有：

$$\begin{aligned}E&=\int_{r_0}^{r}\left\{\sum_{i=1}^{M}a_i(\xi_0)u_i(\xi)-H(\xi-\xi_0)\right\}^2\mathrm{d}\xi\\&=\sum_{i=1}^{M}\sum_{j=1}^{M}a_i(\xi_0)a_j(\xi_0)\int_{r_0}^{r}u_i(\xi)u_j(\xi)\mathrm{d}\xi-\\&\quad 2\sum_{i=1}^{M}a_i(\xi_0)\int_{r_0}^{r}u_i(\xi)\mathrm{d}\xi+(r-\xi_0)\end{aligned} \tag{4.92}$$

令

$$\begin{aligned}&G_{ij}=(u_i(\xi),u_j(\xi))\\&b_i(\xi_0)=2\int_{\xi_0}^{r}u_j(\xi)\mathrm{d}\xi\end{aligned} \tag{4.93}$$

则(4.92)式可化为：

$$E=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}+(r-\xi_0)$$

式中：$\boldsymbol{A}$ 如(4.84)式所示；$\boldsymbol{B}$ 的要素如(4.93)式所示。

根据最优化原理，求 $\boldsymbol{E}$ 相对于 $\boldsymbol{A}^{\mathrm{T}}$ 的偏导数，并设其为零，则得：

$$\boldsymbol{G}\boldsymbol{A}=\boldsymbol{B} \tag{4.94}$$

或

$$\boldsymbol{A}=\boldsymbol{G}^{-1}\boldsymbol{B} \tag{4.95}$$

求出向量 $\boldsymbol{A}$ 之后，不难根据(4.79)式和(4.76)式求出平均函数 $A(\xi,\xi_0)$ 和 ξ_0 处模型之平均值 $\langle m(\xi_0)\rangle$。

现在讨论 $\langle m(\xi_0)\rangle$ 之分辨力。如果定义分辨宽度为：

$$W[\delta]=12\int_{r_0}^{r}[\widetilde{H}(\xi-\xi_0)-H(\xi-\xi_0)]^2\mathrm{d}\xi \tag{4.96}$$

显然，$W[\delta]$ 越小，说明 $\widetilde{H}(\xi-\xi_0)$ 越接近 $H(\xi-\xi_0)$，即分辨力越

高；反之，则分辨力越低。为什么取 12 呢？下面将作进一步说明。

假定 $\tilde{\delta}$ 函数（它是 δ 函数的近似函数）为一个中心在 ξ_0 处之方波，其宽度为高度的倒数，且面积为 1，如图 4.7 所示，此时，

$$\tilde{\delta}(\xi,\xi_0)=\begin{cases}\dfrac{1}{a} & (|\xi-\xi_0|\leqslant\dfrac{a}{2})\\ 0 & \text{其他}\end{cases}$$

将上式代入（4.96），得：

$$W[\tilde{\delta}]=a$$

正好和脉冲的宽度 a 相等。如果取（4.96）式中的系数为 1，而不是 12，则 $W[\tilde{\delta}]=a/12$，与实际脉冲的宽度相差一个比例因子 1/12。

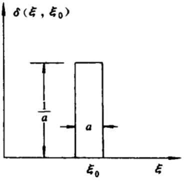


图 4.7 $\tilde{\delta}(\xi-\xi_0)$的一种近似

应特别提出的是，和第一类狄里希来准则类似，按第二类狄里希来准则求得的平均值 $\langle m(\xi_0)\rangle$，也是求在最小方差意义上的解，由于目标函数不同了，所以结果也会发生变化。要知道 $\langle m(\xi_0)\rangle$ 在多大程度上反映 ξ_0 处 $m(\xi)$ 的值，还必须知道平均函数 $A(\xi,\xi_0)$ 的形态（包括其极值和宽度）、分辨力的宽度 $W[\tilde{\delta}]$。只有结合 $\langle m(\xi_0)\rangle$，$A(\xi,\xi_0)$ 和 $W[\tilde{\delta}]$，才有可能给 $\langle m(\xi_0)\rangle$ 接近 $m(\xi_0)$ 的程度作出明确的回答。

（3）BG 展伸准则。为了改善平均函数的形态，提高分辨力，不同的学者提出了各种不同的准则，其中最常用的要算 Backus，G. E 和 Gilbert，F 提出的 BG 展伸（spread）准则了。

BG 展伸准则抛开了 δ 函数或阶跃函数的形态，考虑的是平均函数 $A(\xi,\xi_0)$ 的矩，比如说二阶矩。设目标函数：

$$E=12\int_{r_0}^{r}(\xi-\xi_0)^2A(\xi,\xi_0)^2\mathrm{d}\xi=W[\tilde{\delta}] \tag{4.97}$$

并在 $A(\xi,\xi_0)$ 的规一化条件

$$\int_{r_0}^{r} A(\xi,\xi_0)\mathrm{d}\xi = 1 \tag{4.98}$$

的约束之下求目标函数(4.97)式之极小。在(4.97)式中，$(\xi-\xi_0)^2$ 实际上是权函数。当 $(\xi-\xi_0)\to 0$ 时，权很小；当 ξ 远离 ξ_0 时，权很大。然而，我们欲求的是目标函数 E 之极小，其结果必然是突出了 $\xi\to\xi_0$ 处之平均函数值，使 $A(\xi,\xi_0)$ 更接近于 $\delta(\xi-\xi_0)$ 函数。

根据条件极值原理，可引入拉格朗日算符，将(4.97)式的条件极值问题，化为无条件极值问题。即求泛函

$$E = 12\int_{r_0}^{r}(\xi-\xi_0)^2 A(\xi,\xi_0)^2\mathrm{d}\xi + \lambda\left[1-\int_{r_0}^{r}\sum_{i=1}^{M} a_i(\xi_0)g_i(\xi)\mathrm{d}\xi\right] \tag{4.99}$$

的无条件极值问题。式中：λ 为拉格朗日算符。

若将(4.99)写成

$$E = \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}(\xi_0)a_i(\xi_0)a_j(\xi_0) + \lambda\left[1-\sum_{i=1}^{M}a_i(\xi_0)u_i\right] \tag{4.100}$$

这里，我们利用了(4.79)式和(4.89)式。若设

$$S_{ij}(\xi_0) = 12\int_{r_0}^{r}(\xi-\xi_0)^2 g_i(\xi)g_j(\xi)\mathrm{d}\xi \tag{4.101}$$

并将(4.100)写成矩阵，则有

$$\boldsymbol{E} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{A} + \lambda[1-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}] \tag{4.102}$$

式中：

$$\boldsymbol{S} = \begin{bmatrix} S_{11}(\xi_0) & S_{12}(\xi_0) & \cdots & S_{1M}(\xi_0) \\ S_{21}(\xi_0) & S_{22}(\xi_0) & \cdots & S_{2M}(\xi_0) \\ \vdots & \vdots & & \vdots \\ S_{M1}(\xi_0) & S_{M2}(\xi_0) & \cdots & S_{MM}(\xi_0) \end{bmatrix} \tag{4.103}$$

$$\boldsymbol{U} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix} \tag{4.104}$$

设

$$\frac{\partial E}{\partial \boldsymbol{A}^{\mathrm{T}}}=\boldsymbol{S}\boldsymbol{A}-\lambda\boldsymbol{U}=0$$

则得

$$\boldsymbol{A}=\lambda\boldsymbol{S}^{-1}\boldsymbol{U} \tag{4.105}$$

又因

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}=1 \tag{4.106}$$

所以

$$\lambda\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{U}=1$$

故

$$\lambda=(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{U})^{-1} \tag{4.107}$$

将(4.107)式代入(4.105)式，得：

$$\boldsymbol{A}=(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{U})^{-1}\boldsymbol{S}^{-1}\boldsymbol{U} \tag{4.108}$$

再将(4.108)式代入(4.76)式和(4.79)式，得：

$$\langle m(\xi_0)\rangle=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{d}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{U})^{-1}\boldsymbol{d} \tag{4.109}$$

和

$$A(\xi,\xi_0)=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{g}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{U})^{-1}\boldsymbol{g} \tag{4.110}$$

式中：

$$\boldsymbol{g}=\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix}$$

由(4.108)式可以看出，由于 $\boldsymbol{A}$ 只和核函数有关，所以平均函数 $A(\xi,\xi_0)$ 必然只决定于核函数。而 $\langle m(\xi_0)\rangle$ 不仅与核函数有关，而且还与观测数据 $\boldsymbol{d}$ 有关。从(4.109)式可知，凡是拟合观测数据的模型，都具有相同的平均值 $\langle m(\xi_0)\rangle$。

讨　论：

(1) 按(4.97)式计算的 E，可以近似地表示平均函数的分辨宽

度。当平均函数 $A(\xi,\xi_0)$ 集中在 ξ_0 附近时，其值与平均函数的分辨宽度重合很好。记 $S(\xi_0,A)=E$，则 $S(\xi_0,A)$ 的大小和分辨力成反比。而 $S(\xi_0,A)$ 越小，则分辨力越高。

(2) 由 BG 展伸准则求出的平均函数的宽度，一般比狄里希来准则求出的平均函数的要宽，但其边叶要小，不存在负值，则相应的 $\langle m(\xi_0)\rangle$ 的方差较小。由于平均函数的边叶不出现负值，因而给 $\langle m(\xi_0)\rangle$ 的解释带来了方便。

(3) 基于 BG 展伸准则工作时，由于每一个 ξ_0 的加权系数都不相同，从而大大增加了计算内积矩阵(4.101)式和解方程(4.108)式之工作量，从而增加了计算成本。

(4)定义 E 时，应加一系数 12。

如取 $A(\xi,\xi_0)$ 为中心在 ξ_0 处之单位矩形脉冲，如图 4.8 所示。即

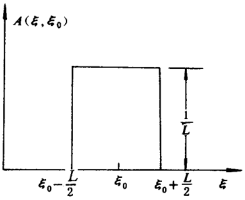


图 4.8　展伸准则的一种近似

$$A(\xi,\xi_0)=\begin{cases} L^{-1} & -\dfrac{L}{2}\leqslant\xi-\xi_0 \\ L^{-1} & \dfrac{L}{2}\geqslant\xi-\xi_0 \\ 0 & \text{其他位置} \end{cases}$$

将上述定义代入(4.97)式得：

$$
\begin{aligned}
E &= S(\xi_0, A) \\
&= 12\int_{r_0}^{r} (\xi-\xi_0)^2 A(\xi,\xi_0)^2 \mathrm{d}\xi \\
&= 12\int_{-\frac{L}{2}}^{\frac{L}{2}} x^2 A(x_1,\xi)^2 \mathrm{d}x \\
&= 12 \cdot \frac{1}{3}x^2 \cdot \frac{1}{L^2}\Big|_{-L/2}^{L/2} = L
\end{aligned}
$$

由此可见，在这种条件下，$S(\xi_0,A)$恰好等于矩形的宽度。在实践中，$S(\xi_0,A)$可作为平均函数 $A(\xi,\xi_0)$半极值之宽度，即作为分辨力的量度。

综上面所述，BG线性评价的信息包括在三个量中，即平均值$\langle m(\xi_0)\rangle$、平均函数 $A(\xi,\xi_0)$和分辨力 $S(\xi_0,A)$。只有综合分析以上三种信息才可能取得关于 ξ_0 处地球物理模型的真正信息。

§4 BG 线性评价(二)

在上一节中，讨论了在观测数据为有限、精确的情况下，BG 的线性评价问题。可以看出，虽然反演问题的解是非唯一的，但在窗口$A(\xi,\xi_0)$范围内，模型 $m(\xi)$之平均值$\langle m(\xi_0)\rangle$却是唯一的。这就表明，不管用什么方法求出来的解，只要能拟合观测数据，都具有相同的平均值(即模型与平均函数的内积)。

然而，实际观测数据都具有误差。在有误差的观测数据时，如何进行线性评价？前一节中所讨论的 BG 线性评价理论是否仍然适用？现在，我们将来回答这个问题。

设

$$\tilde{d}_i = d_i' + \delta\tilde{d}_i \qquad (i=1,2,\cdots,M) \qquad (4.111)$$

式中：$\tilde{d}_i$，$\delta\tilde{d}_i$ 分别为第 i 个观测数据和方差；而 d_i' 是第 i 个数据的真值。

利用(4.78)式，讨论具有误差的 ξ_0 处的模型之平均值，则有：

$$\langle \widetilde{m}(\xi_0) \rangle = \sum_{i=1}^{M} a_i(\xi_0)\widetilde{d}_i$$

$$= \sum_{i=1}^{M} a_i(\xi_0) d_i' + \sum_{i=1}^{M} a_i(\xi_0)\delta\widetilde{d}_i$$

将$\langle \widetilde{m}(\xi_0) \rangle$记为：

$$\langle \widetilde{m}(\xi_0) \rangle = \langle m(\xi_0) \rangle + \langle \delta\widetilde{m}(\xi_0) \rangle \tag{4.112}$$

式中：$\langle m(\xi_0) \rangle$是平均值之真值；$\langle \delta\widetilde{m}(\xi_0) \rangle$是$\xi_0$处平均值之误差。

为了定量估计观测数据的误差$\delta\widetilde{d}_i(i=1,2,\cdots,M)$对反演结果和评价结果的影响，必须了解观测数据误差所遵循的规律。这里，不妨假定观测数据误差$\delta\widetilde{d}_i(i=1,2,\cdots,M)$服从高斯分布、零平均值且方差为$\sigma_i(i=1,2,\cdots,M)$的统计特性，并设：

$$E[\delta\widetilde{d}_i]=0$$

为数学期望，而

$$E_{ij}=\mathrm{cov}[\delta\widetilde{d}_i,\delta\widetilde{d}_i] \tag{4.113}$$

为协方差矩阵。

假定$\delta\widetilde{d}_i$是独立的，则有：

$$E_{ij}=\begin{cases}\sigma_i^2 & (i=j)\\ 0 & (i\neq j)\end{cases} \tag{4.114}$$

由(4.112)式知：

$$E[\langle \widetilde{m}(\xi_0) \rangle]=E[\langle m(\xi_0) \rangle]+E[\langle \delta\widetilde{m}(\xi_0) \rangle]$$

$$=\langle m(\xi_0) \rangle + \sum_{i=1}^{M} a_i(\xi) E[\delta\widetilde{d}_i]$$

$$=\langle m(\xi_0) \rangle$$

也就是说，虽然观测值有误差，但只要满足高斯分布、零平均值的统计特性，此时$E[\langle \widetilde{m}(\xi_0) \rangle]$仍与平均值$\langle m(\xi_0) \rangle$完全一样。

而$\langle \widetilde{m}(\xi_0) \rangle$的方差为：

$$\mathrm{Var}[\langle m(\xi_0) \rangle]=\mathrm{Var}[\langle \delta\widetilde{m}(\xi_0) \rangle]$$

$$= \mathrm{Var}[\sum_{i=1}^{M} a_i(\xi_0)\hat{\delta}d_i]$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M} a_i(\xi_0)a_j(\xi_0)E_{ij}$$

$$= \sum_{i=1}^{M} a_i^2(\xi_0)\sigma_i^2$$

或

$$V^2(\xi_0) = \sum_{i=1}^{M} a_i^2(\xi_0)\sigma_i^2 \tag{4.115}$$

由此可见，在 ξ_0 处模型的平均值 $\langle m(\xi_0)\rangle$ 之方差 $V^2(\xi_0)$，不仅与观测数据的方差 σ_i^2 有关，而且与待定系数 $a_i(\xi_0)(i=1,2,\cdots,M)$ 有关。因此，在选择计算 $a_i(\xi_0)$ 的方法时，既要考虑分辨力 $S(\xi_0,A)$，也应考虑方差 $V^2(\xi_0)$，以避免分辨力很高但方差也很大的现象出现。

当观测数据有误差时，从统计观点看，分辨力和方差是矛盾的。分辨力高(即 $S(\xi_0,A)$ 小)，方差大；分辨力低，方差小。结论是，必须在方差 $V^2(\xi_0)$ 和分辨力 $S(\xi_0 A)$ 之间取折衷，并以此作为求取待定系数 $a_i(\xi_0)$ 之准则。待 $a_i(\xi_0)$ 求出之后，再按(4.115)式和(4.97)式计算 $V^2(\xi_0)$ 和 $S(\xi_0,A)$ 以及模型在 ξ_0 处之平均值 $\langle m(\xi_0)\rangle$。

1. 折衷准则

在观测数据有误差的情况下，对反演问题提出既要分辨力高，又要方差小的要求是不现实的，也是不可能的。

考虑

$$E = W[\delta]\cos\theta + V^2(\xi_0)\sin\theta \tag{4.116}$$

式中：E 为目标函数；θ 为折衷参数；$\cos\theta$、$\sin\theta$ 分别表示对分辨力 $W[\delta]$ 和方差 $V^2(\xi_0)$ 的加权因子(加权因子满足 $\sin^2\theta+\cos^2\theta=1$)。当 θ 加大时，方差对目标函数的贡献加大，极小 E 求得的 $\langle m(\xi_0)\rangle$ 方差减小，精度提高，但分辨力降低，因此，当 $\theta=\frac{\pi}{2}$ 时，$\langle m(\xi_0)\rangle$ 的精度最高，分辨力最低；相反，θ 减小时，解的分辨力提高，方差增大，精度降

低。当 $\theta=0$ 时，分辨力最高，方差也最大。

现在，我们讨论如何实现 BG 线性评价。

将(4.97)式和(4.115)式中的 $W[\delta]$ 和 $V_2(\xi_0)$ 代入(4.116)式，并取 $\sigma_i=1$，即设观测数据都为单位方差，则得目标函数。

$$E=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{A}\cos\theta+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\sin\theta \tag{4.117}$$

欲在

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}=1$$

条件限制之下，求(4.117)式之极小。和以前一样，这种条件极值问题，可以引入拉格朗日算符 λ 化为无条件极值问题，此时，(4.117)式变为：

$$E=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{A}\cos\theta+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\sin\theta-\lambda\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U} \tag{4.118}$$

求 E 相对于 $\boldsymbol{A}^{\mathrm{T}}$ 之导数，并设

$$\frac{\partial E}{\partial \boldsymbol{A}^{\mathrm{T}}}=0$$

则得

$$\lambda\boldsymbol{U}=\boldsymbol{S}\boldsymbol{A}\cos\theta+\boldsymbol{A}\sin\theta \tag{4.119}$$

设

$$\boldsymbol{S}=\boldsymbol{O}\boldsymbol{\Lambda}\boldsymbol{O}^{\mathrm{T}}$$

式中：Λ 为对称矩阵 S 之特征值组成的对角线矩阵；$\boldsymbol{O}$ 为 S 之特征向量组成的特征向量矩阵。

若以 $\underline{x}$ 表示向量 x 按 $\boldsymbol{O}^{\mathrm{T}}$ 转轴后之转轴向量，则

$$\underline{\boldsymbol{A}}=\boldsymbol{O}^{\mathrm{T}}\boldsymbol{A},\qquad \underline{\boldsymbol{U}}=\boldsymbol{O}^{\mathrm{T}}\boldsymbol{U} \tag{4.120}$$

对(4.119)式作简单运算后，可写为：

$$\lambda\underline{\boldsymbol{U}}=\Lambda\underline{\boldsymbol{A}}\cos\theta+\underline{\boldsymbol{A}}\sin\theta=\cos\theta\Lambda\underline{\boldsymbol{A}}+\sin\theta\underline{\boldsymbol{A}}$$

因而有：

$$\underline{\boldsymbol{A}}=\lambda(\cos\theta\cdot\Lambda+\sin\theta\cdot\boldsymbol{I})^{-1}\underline{\boldsymbol{U}} \tag{4.121}$$

又因

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{O}\boldsymbol{O}^{\mathrm{T}}\boldsymbol{U}=\underline{\boldsymbol{A}}^{\mathrm{T}}\underline{\boldsymbol{U}}=\underline{\boldsymbol{U}}^{\mathrm{T}}\underline{\boldsymbol{A}}=\boldsymbol{I} \tag{4.122}$$

并将(4.121)式代入(4.122),

$$\lambda \underline{\boldsymbol{U}}^{\mathrm{T}}(\cos\theta\Lambda+\sin\theta\cdot\boldsymbol{I})^{-1}\underline{\boldsymbol{U}}=\boldsymbol{I}$$

故得:

$$\lambda \boldsymbol{I}=(\underline{\boldsymbol{U}}^{\mathrm{T}}(\cos\theta\cdot\Lambda+\sin\theta\boldsymbol{I})\underline{\boldsymbol{U}})^{-1} \tag{4.123}$$

将(4.123)代入(4.121)得$\underline{\boldsymbol{A}}$。求得$\underline{\boldsymbol{A}}$后,就不难求展伸系数$S(\xi_0,A)$及方差$V^2(\xi_0)$,

$$S(\xi_0,A)=\underline{\boldsymbol{A}}^{\mathrm{T}}\Lambda\underline{\boldsymbol{A}} \tag{4.124}$$

$$V^2(\xi_0)=\underline{\boldsymbol{A}}^{\mathrm{T}}\underline{\boldsymbol{A}} \tag{4.125}$$

而平均值

$$\langle m(\xi_0)\rangle=\int_{r_0}^{r}A(\xi,\xi_0)m(\xi)\mathrm{d}\xi$$

$$=\sum_{i=1}^{M}a_i(\xi_0)\hat{d}_i=\boldsymbol{A}^{\mathrm{T}}\tilde{\boldsymbol{d}}=\underline{\boldsymbol{A}}^{\mathrm{T}}\underline{\tilde{\boldsymbol{d}}} \tag{4.126}$$

其中

$$\underline{\tilde{\boldsymbol{d}}}=\boldsymbol{O}^{\mathrm{T}}\tilde{\boldsymbol{d}} \tag{4.127}$$

从(4.124)式~(4.127)式可以看出,由于$\boldsymbol{A}$或$\underline{\boldsymbol{A}}$与折衷参数$\theta$有关,所以$S(\xi_0,A)$,$V^2(\xi_0)$以及$\langle m(\xi_0)\rangle$都随$\theta$而变化。当$\theta=0$时,$S(\xi_0,A)$最小,分辨力最高,但方差最大;当$\theta=\frac{\pi}{2}$时,此时分辨力最低,方差最小。计算时,必须根据观测数据方差的大小,选择最佳的θ值,以使方差$V^2(\xi_0)$和分辨力$S(\xi_0,A)$合符要求。

2. 折衷曲线

如上所述,为了求得模型$m(\xi)$在ξ_0处的平均值$\langle m(\xi_0)\rangle$,必须选择不同的折衷参数$\theta_i(0\leqslant\theta_i\leqslant\frac{\pi}{2})$,并计算相应于$\theta_i$之如下参数:

(1)方差$V_i^2(\xi_0)$;

(2)展伸系数$S_i(\xi_0,A)$(对第一、第二类狄里希来准则而言,为$W_i[\delta]$,即分辨宽度)。

然后,根据不同θ_i计算的$V_i^2(\xi_0)$和$S_i(\xi_0,A)$画出一条曲线,即

折衷曲线，如图 4.9 所示。再根据折衷曲线选择最佳的折衷参数 θ_i，计算最佳的 $a_i(\xi)$ 及相应的平均函数 $A_i(\xi,\xi_0)$ 和平均值 $\langle m_i(\xi_0)\rangle$，并以此作为 ξ_0 处反演结果之最后取值。

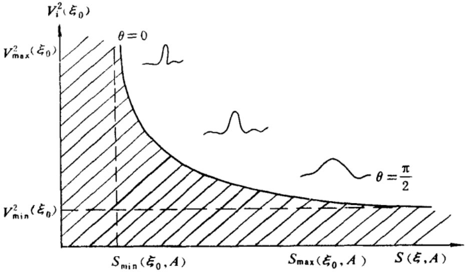


图 4.9　折衷曲线的形态

之后，计算另一 ξ_0 处之折衷曲线和平均函数 $A(\xi,\xi_0)$ 和平均值 $\langle m(\xi_0)\rangle$。

同一反演问题不同 ξ_0 处所需的最佳折衷参数和绘出的折衷曲线不可能完全相同。一般说来，ξ_0 小时（浅处）折衷曲线随折衷参数变化明显，最佳折衷参数易于确定，如图 4.10 中 A 所示；随着深度的加大，折衷曲线趋于平缓，最佳折衷参数难予确定，如图 4.10 中 C 所示。

值得提出的是，不同地球物理反演问题，折衷曲线的形态也不相同。分辨力高的地球物理方法折衷曲线变化明显，距原点较近；相反，分辨力低的地球物理勘探方法的折衷曲线比较平缓，距原点较远。

折衷曲线的形态完全取决于核函数的性质，取决于地球物理问

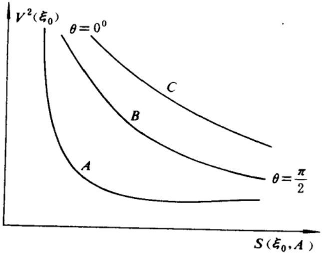


图 4.10　折衷曲线的类型

题的物理本质。一般说来，波场勘探，折衷曲线较佳，呈现出 A，B 类；而位场勘探，折衷曲线较差，为 B、C 类。

当然，折衷曲线还和观测数据的方差有关，方差越大，折衷曲线越趋于平缓。

除了上面一些原因外，折衷曲线还和观测资料的多少有关。M 增加，分辨力应该提高，相应折衷曲线向原点移动，如图 4.11 所示。当然，不是 M 越大，效果越佳。事实说明，对一定地球物理问题而言，M 增大到一定程度后，对分辨力就不再起明显的作用。况且，M 加大，野外采集工作量增加，成本提高。因此，在实际工作中，不能一味增加观测数据的点数 M，必须结合分辨力，地质任务、成本等诸多因素进行综合考虑。实际上，在一个地区施工之前，就应该在一定假设前提下，绘出不同 M 的折衷曲线，以确定最佳观测值的个数，此即实验设计。

既然，在求取 $\langle m(\xi_0)\rangle$ 的同时，还得到了估计 ξ_0 处模型平均值 $\langle m(\xi_0)\rangle$ 之方差和分辨宽度这些反映解的质量之重要参数。因此，在

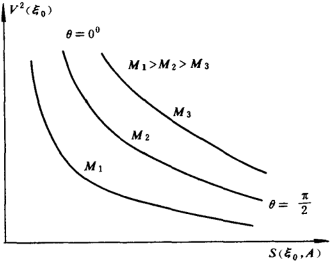


图 4.11 折衷曲线随 M 的变化

表示 $\langle m(\xi_0)\rangle$ 时，最好把它们也同时表示出来。如图 4.12 所示，垂线表示方差，水平线表示分辨宽度，其交点之坐标就是 $\langle m(\xi_0)\rangle$。看了图 4.12，对反演的结果及其质量评价就一目了然了。

3. 所谓的平均模型

在 BG 线性评价理论中已经证明，$\langle m(\xi_0)\rangle$ 是从观测数据中可以提取出来的地球模型之唯一信息。试问，不同 ξ_i 的平均值 $\langle m(\xi_i)\rangle$ 所构成的"平均模型"是否可作为我们反演构制的地球物理模型？是否能拟合观测数据呢？回答一般是否定的，因为，可以证明，除了：

(1)折衷参数 $\theta=0$；

(2)没有平均函数为单位模这一限制条件的第一狄里希来准则外，其他任何准则所计算的"平均模型"都不可能重建观测数据或拟合观测数据。

为说明以上结论，让我们研究满足以上两个条件的第一类狄里希来问题。此时，目标函数为：

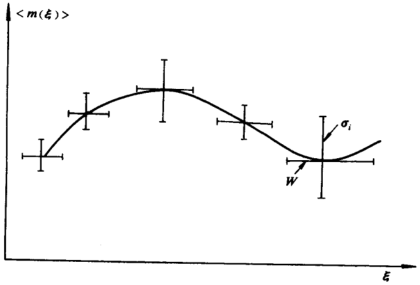


图 4.12 $\langle m(\xi_0)\rangle$的图示法

$$E=\int_{r_0}^{r}[A(\xi,\xi_0)-\delta(\xi-\xi_0)]^2\mathrm{d}\xi \tag{4.128}$$

将(4.79)式代入上式，则有：

$$E=\int_{r_0}^{r}[\sum_{i=1}^{M}a_i(\xi_0)g_i(\xi)-\delta(\xi-\xi_0)]^2\mathrm{d}\xi$$

如采用(4.84)式、(4.85)式的符号，可得：

$$\boldsymbol{A}=\boldsymbol{G}^{-1}\boldsymbol{d}$$

且

$$\langle m(\xi_0)\rangle=\sum_{j=1}^{M}a_j(\xi_0)d_j$$

现在，将$\langle m(\xi_0)\rangle$作为模型代入数据方程，即有：

$$\begin{aligned}\tilde{d}_l&=\int_{r_0}^{r}\langle m(\xi)\rangle g_l(\xi)\mathrm{d}\xi\\&=\int_{r_0}^{r}\sum_{j=1}^{M}a_j(\xi)d_j\cdot g_l(\xi)\mathrm{d}\xi\end{aligned}$$

$$
\begin{aligned}
&=\int_{r_0}^{r}\sum_{j=1}^{M}\sum_{k=1}^{M}(\boldsymbol{G}^{-1})_{jk}g_k(\xi)d_jg_l(\xi)\mathrm{d}\xi \\
&=\sum_{j=1}^{M}\sum_{k=1}^{M}d_j(\boldsymbol{G}^{-1})_{jk}\int_{r_0}^{r}g_l(\xi)g_k(\xi)\mathrm{d}\xi \\
&=\sum_{j=1}^{M}\sum_{k=1}^{M}d_j(\boldsymbol{G}^{-1})_{jk}\boldsymbol{G}_{kl} \\
&=\sum_{j=1}^{M}d_j\delta_{jl} \\
&=d_l
\end{aligned}
$$

可见，只要折衷参数 $\theta=0$，则(4.116)式在第一类狄里希来准则下，就简化为(4.128)式。当只有极小(4.128)式时，没有其他任何限制条件，才会满足上述两个条件。以上证明表示，这样得到的“平均模型”就完全拟合观测数据（事实上，这时的“平均模型”就是最小模型）。除此以外，在其他任何准则下，所获得的“平均模型”都不拟合观测数据，故不能作为可接受的模型。

从以上的讨论可以看出，BG 线性反演理论中的模型构制和评价，是既有联系，又不相同的两个独立的问题。它们的任务和在反演理论中的作用也不相同。前者是在一定限制条件下，寻找一个解释人员认为可以接受的、最优的解。这种最优是有限制条件的，并不一定是在接近真实地球物理模型意义上的最优。换言之，不同的标准，不同的限制条件以及不同的模型构制方法都可能得到不同的最优解；反演问题的解无论如何是非唯一的。而 BG 的线性评价，就是在承认解是非唯一的前提下，欲从任何一个与真实地球物理模型线性接近的、能拟合观测数据的模型中，提取它与真实地球物理模型之间存在的共同的，也即从观测数据中所能提取出来的唯一信息。也就是对构制的模型进行评价。可见，模型构制和评价是反演过程中的两个不同阶段，各自具有不同的任务和作用。前者是后者的基础和前提，后者是前者的继续和深化。两者互相联系，缺一不可；两者相结合才构成反演问题的整体。那种以为构制出一种模型就完成了反演问题的全

部任务，就万事大吉了，显然是不正确的。

4. 实例

图 4.13 是本章 § 2“例一”的评价结果。图中 a 是 $x=0.05$，0.25，0.5 三个深度处的平均曲线；b～d 分别是相应深度之平均曲线，其中，曲线 1，2，3 分别表示平均曲线的方差为 0.01，0.05，0.1。

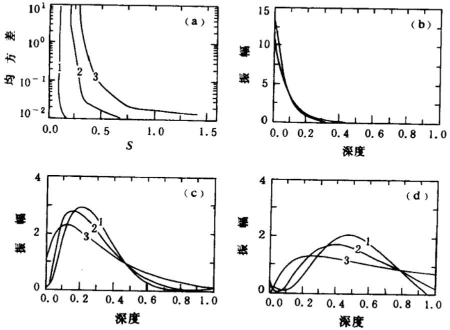


图 4.13 理论数据的折衷曲线和平均函数

图 a 说明，x 越小，折衷曲线越靠近原点，折衷曲线变化越剧烈，牺牲一点分辨力可换来方差的大幅度降低。随着 x 的加大，折衷曲线越来越平缓，最佳折衷参数越难确定；从图 b 可以看出，不管是方差为 0.01，0.05 还是 0.1，平均函数的极值都集中在 0.05 附近，分辨力都很高。当 $x=0.25$ 时，平均函数（图 c）的极值只有当 $\sigma_0=0.01$

时，才集中在 $x=0.25$ 处，当 $\sigma_0=0.05, 0.1$ 时，极值点均向小 x 方向移动，且平均函数的半极值宽度都超过 0.25。此时之分辨力是相当低的。当 $x=0.05$ 的图 c 时，平均函数的性质变得更为恶劣，分辨力更低。因此，在这样一些平均函数范围内，虽然 $\langle m(x)\rangle$ 是唯一的，但这种唯一性对提取真实地球物理模型的信息而言，不会具有更多的价值。

平均函数的分辨力，完全由核函数的性质所决定的。就是同一个地球物理问题，采集方式不同，平均函数的性质也不尽相同。一般说来，基于波场勘探的地球物理问题，比基于位场勘探的问题分辨力要高，而且深度越大，平均函数的分辨力越低，这是不以人们意志为转移的客观规律。

图 4.14 是标准方差分别为 0.01，0.05，0.1 和 0.5 时，上述例子

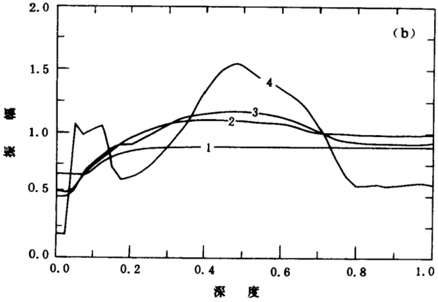


图 4.14 理论数据的 $\langle m(x)\rangle$-x 曲线

之平均模型，即$\langle m(x)\rangle$-x曲线。从图中的四条曲线可以看出，它们和真实模型(图4.2)均相距甚远。以这种平均模型作为模型无论如何也不可能拟合观测数据的。

§5 BG反演理论在反褶积中的应用

褶积和反褶积是在线性系统中经常遇到的问题，在地球物理学中具有重要意义。计算反褶积的方法很多，这里我们研究一下把反褶积看成反演问题，利用BG反演理论作反褶积的途径、方法和问题。

设$W(t)$代表线性系统的响应函数，比如说地震子波；$m(t)$代表地球物理模型，比如说地震波的反射函数；则$x(t)$代表系统的输出，比如说地震记录，在无噪声的情况下，有

$$\begin{aligned} x(t) &= m(t) * w(t) \\ &= \int_0^{t_A} m(\tau) w(t-\tau) \mathrm{d}t \end{aligned} \tag{4.129}$$

式中：*表示褶积；而$(0,t_A)$是$m(t)$的定义域。
则有：

$$\begin{aligned} x_j = x(t_j) &= \int_0^{t_A} m(\tau) w(t_j-\tau) \mathrm{d}\tau \\ &= \int_0^{t_A} m(\tau) g_j(\tau) \mathrm{d}\tau \end{aligned} \tag{4.130}$$

式中：$g_j(\tau)=w(t_j-\tau)$是第j个观测数据x_j所对应的子波。在反演理论中，常将$g_j(\tau)$称为核函数；$m(\tau)$称为模型；把x_j称为观测数据。

如果核函数$g_j(\tau)$是一个δ函数，观测数据x_j就是t_j处模型$m(\tau)$之值。否则，x_j将是模型$m(t)$在窗口$g_i(\tau)$范围内的加权平均。

根据BG线性评价理论，在t_0处模型的平均值应是

$$\begin{aligned} \langle m(t_0)\rangle &= \sum_{j=1}^{M} a_j(t_0) x_j \\ &= \int_0^{t_A} m(t) A(t,t_0) \mathrm{d}t \end{aligned} \tag{4.131}$$

而

$$A(t,t_0)=\sum_{j=1}^{M}a_j(t_0)g_j(t) \tag{4.132}$$

式中：$A(t,t_0)$称为平均函数；$a_j(t_0)$是一组待定系数；$\langle m(t_0)\rangle$是我们可以从观测数据中获得的模型在 t_0 处的唯一信息。

正如前面所指出的，由有限个函数 $g_j(t)$的线性组合，是不可能构成一个 $\delta(t,t_0)$函数，只能是尽可能地逼近。由“逼近”的准则不同，则平均函数 $A(t,t_0)$就不一样。如用第一类狄里希来准则，则极小应为

$$E(t_0)=\int_0^{t_A}[A(t,t_0)-\delta(t-t_0)]^2\mathrm{d}t \tag{4.133}$$

值得注意的是，在地震勘探中，可以近似地认为子波 $w(t)$不随时间传播而变化。因此，任何时刻 t_i 的子波都是完全相同的。这就给 BG 反演带来了极大的方便。比如，在极小 $E(t_0)$时，就没有必要加以下限制条件，即

$$\int_0^{t_A}A(t,t_0)\mathrm{d}t=1$$

显然，这种限制条件在某些情况下是十分重要的。

$E(t_0)$之极小值，不难理解是 $A(t,t_0)$接近 $\delta(t-t_0)$程度的一种量度。故也称

$$W=\frac{1}{A(t,t_0)} \tag{4.134}$$

为分辨力。

在观测数据 x_j 具有误差时，情况就不同了。假设 x_j 之误差为 $n_j(t)$，且服从均值为零，方差为 σ_j 的高斯分布。则$\langle m(t_0)\rangle$之方差为：

$$V^2(t_0)=\mathrm{Var}[\langle m(t_0)\rangle]=\sum_{i=1}^{M}\sum_{j=1}^{M}a_iE_{ij}a_j \tag{4.135}$$

式中：

$$E_{ij}=\mathrm{cov}[n(t_i)n(t_j)]$$

是观测数据协方差矩阵 E 之要素。

如前所述，我们必须在方差和分辨力之间取折衷，使目标函数变

为：

$$\overline{E}(t_0)=E(t_0)\cos\theta+V^2(t_0)\sin\theta \tag{4.136}$$

这里 θ 是折衷参数，且 $0\leqslant\theta\leqslant\frac{\pi}{2}$。

将(4.133)式、(4.135)式代入(4.136)式并利用(4.132)式求解。设

$$\frac{\partial\overline{E}(t_0)}{\partial \boldsymbol{A}}=0$$

简化后，得：

$$\cos\theta[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}-\boldsymbol{r}]+\sin\theta\cdot\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}=0 \tag{4.137}$$

式中：

$$\boldsymbol{A}=\begin{bmatrix}a_1(t_0)\\a_2(t_0)\\\vdots\\a_M(t_0)\end{bmatrix},\qquad \boldsymbol{r}=\begin{bmatrix}r_1\\r_2\\\vdots\\r_M\end{bmatrix}$$

而

$$r_k=\int_0^{t_A}g_k(t)\delta(t-t_0)\mathrm{d}t \qquad (k=1,2,\cdots,M)$$

$$\boldsymbol{G}=\begin{bmatrix}G_{11}&G_{12}&\cdots&G_{1M}\\G_{21}&G_{22}&\cdots&G_{2M}\\\vdots&\vdots&&\vdots\\G_{M1}&G_{M2}&\cdots&G_{MM}\end{bmatrix}$$

同样：

$$G_{jk}=\int_0^{t_A}g_j(t)g_k(t)\mathrm{d}t \qquad (j=1,2,\cdots,M)$$

$$\boldsymbol{E}=\begin{bmatrix}E_{11}&E_{12}&\cdots&E_{1M}\\E_{21}&E_{22}&\cdots&E_{2M}\\\vdots&\vdots&&\vdots\\E_{M1}&E_{M2}&\cdots&E_{MM}\end{bmatrix}$$

这里 $\boldsymbol{E}$ 是协方差矩阵。现在我们讨论(4.137)式的解法：

令 $\boldsymbol{G}=\boldsymbol{O}\Lambda\boldsymbol{O}^{\mathrm{T}}$，定义转轴后的向量 $\boldsymbol{A}$ 和 $\boldsymbol{r}$ 为：

$$\underline{\boldsymbol{A}}=\boldsymbol{O}^{\mathrm{T}}\boldsymbol{A},\qquad \underline{\boldsymbol{r}}=\boldsymbol{O}^{\mathrm{T}}\boldsymbol{r} \tag{4.138}$$

且令

$$\boldsymbol{D}=\cos\theta\cdot\Lambda+\sin\theta\cdot E$$

在 $E=\sigma_0^2\boldsymbol{I}$ 时，

$$\boldsymbol{D}=\cos\theta\cdot\Lambda+\sin\theta\cdot\sigma_0^2\boldsymbol{I}$$

则(4.137)式的解为：

$$\underline{\boldsymbol{A}}^{\mathrm{T}}=\underline{\boldsymbol{r}}^{\mathrm{T}}\cdot\boldsymbol{D}^{-1}\cos\theta \tag{4.139}$$

在求得 $\underline{\boldsymbol{A}}$ 后，可按(4.138)式直接计算 $\boldsymbol{A}$，进而求出平均函数 $A(t,t_0)$ 及 $\langle m(t_0)\rangle$。

现在，我们来分析精确数据和加有 5%白噪声的理论数据——人工合成记录，按 BG 反演方法求得的解。

图 4.15(a)是合成地震记录用的反射系数；图 15(b)的右边是子波，左边是人工合成记录，合成时按(4.130)进行褶积运算，取 $N=41$，采样间隔 $\Delta=0.0125$ 秒，$t_A=1.0$ 秒；图 4.15(c)左边的曲线是 $\langle m(t)\rangle$，右边是 $\theta=0$ 时之平均函数。可以看出，在精确观测数据的情况下，平均模型 $\langle m(t)\rangle$ 与真实模型，即波阻抗介面的反射系数几乎一一对应。

图 4.16(a)是反射系数；(b)的右边是子波，左边是加有 5%白噪声的人工合成记录；(c)图的右边平均函数 $A(\xi,\xi_0)$，左边是 $\langle m(t)\rangle$，显然 $\langle m(t)\rangle$ 的分辨力很高，方差很大(SIG=140)，由于方差太大，从 $\langle m(t)\rangle$ 图上得不出关于模型的任何信息；随着平均函数分辨力的降低，方差的减小，$\langle m(t)\rangle$ 越来越能反映真实模型 $m(t)$ 的形态。直到 4.16(f)，即 $w=0.52$，SIG=0.70 时，从 $\langle m(t)\rangle$ 中，可以得到其模型 $m(t)$ 的主要信息。

前面已讲述了在时间域进行反褶积的方法，根据褶积定理，时域上的褶积等于频率域上的乘积。如以小写字母表示时间域的函数，大写字母表示其谱，则下式成立，

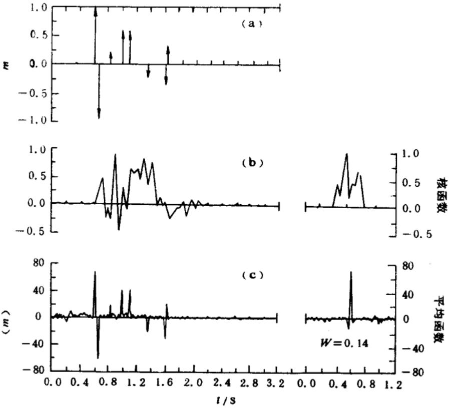


图 4.15　精确数据人工合成地震记录，及其 BG 反演的结果

$$x_j = x(t_j) = \int_0^{t_A} m(t) w(t_j - t) \mathrm{d}t$$

$$\underline{\overline{X}}_k = \underline{\overline{X}}(f_k) = TM_k W_k \tag{4.140}$$

式中：$\underline{\overline{X}}_k$，$M_k$，$W_k$ 分别是周期为 T 的时间序列 $x(t)$、$m(t)$ 和 $w(t)$ 之傅氏变换，由于 w_k 是一个带限记录，所以 $\underline{\overline{X}}_k$ 必然也是带限的。换言

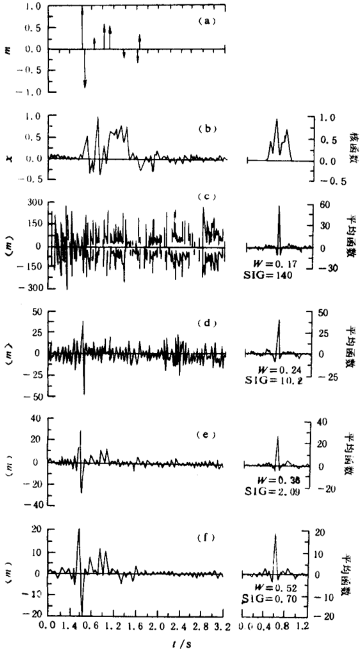


图 4.16　有误差的人工合成地震记录，及其 BG 反演结果

之，当 $f_k > f_c$（f_c 为最小带限频率）时 $w(f)$ 的能量，因而 $\overline{\underline{X}}(f_k)$ 的能量必为零。如 $f_k > f_c$ 时，$\overline{\underline{X}}(f_k)$ 的振幅不为零，则必然是来自于噪声。当然，在反褶积所求得的模型 $M(f)$ 中，也不应含有（$>f_c$）的高频信息。

下面，我们研究一下如何在频率域上实现反褶积。为了便于理解，我们还必须从时域讲起。设：

$$x(t)=m(t)*w(t)+n(t) \tag{4.141}$$

假定：$x(t)$、$m(t)$，$w(t)$ 和 $n(t)$ 都是周期为 T 的周期函数。与（4.79）式相应，在连续介质情况下，设 $a(t,t_0)$ 对应于 $A(t,t_0)$，$w(t)$ 对应于 $g(t)$，$v(t,t_0)$ 对应于 $a(t_0)$，我们的目的是求一连续函数 $v(t)$，以使平均函数。

$$a(t,t_0)=w(t)*v(t_0) \tag{4.142}$$

尽可能接近于 δ 函数。

对（4.141）两端同进行褶积运算，并利用（4.142）式

$$x(t)*v(t_0)=m(t)*a(t,t_0)+n(t)*v(t_0)$$

定义：

$$\langle m(t_0)\rangle=m(t)*a(t,t_0)$$

$$=m(t)*w(t)*v(t_0)$$

$$\langle m(t_0)\rangle=x(t)*v(t_0)$$

$$\delta\langle m(t_0)\rangle=n(t)*v(t_0)$$

故

$$\langle m(t_0)\rangle=\langle m(t)\rangle^{+}+\delta\langle m(t_0)\rangle \tag{4.143}$$

式中：$\delta\langle m(t_0)\rangle$ 是 $\langle m(t)\rangle$ 的方差；$\langle m(t)\rangle^{+}$ 是真值。

遵照以前的方法，设目标函数

$$E=\cos\theta\int_{-\frac{T}{2}}^{\frac{T}{2}}|a(t,t_0)-\delta(t)|^2\mathrm{d}t+\sin\theta\cdot\mathrm{Var}[\delta\langle m(t_0)\rangle] \tag{4.144}$$

假定噪声是平稳的，则方差项可写为

$$\mathrm{Var}[\delta\langle m(t_0)\rangle]=\mathrm{Var}[n(t)*v(t_0)]$$

根据 Parseval 定理，则有：

$$\mathrm{Var}[\delta\langle m(t_0)\rangle]=T^2\sum_{k=-\infty}^{\infty}|N_kV_k|^2 \tag{4.145}$$

把 Parseval 定理应用于(4.144)式的第一项，并利用(4.145)式则得：

$$E=\cos\theta\cdot T\sum_{k=-\infty}^{\infty}\left|TW_kV_k-\frac{1}{T}\right|^2+\sin\theta\cdot\sum_{k=-\infty}^{\infty}|TN_kV_k|^2 \tag{4.146}$$

如以 $\overline{V}_k$，$\overline{W}_k$ 表示 V_k、W_k 之共轨，并设

$$\frac{\partial E}{\partial\overline{V}_k}=0$$

则得：

$$V_k=\frac{\cos\theta\cdot\overline{W}_k}{T^2|W_k|^2\cos\theta+T|N_k|^2\sin\theta}\qquad(k=0,\pm1,\pm2,\pm3,\cdots) \tag{4.147}$$

由于 W_k 是带限的，当 $|f_k|>f_c$ 时，$V_k=0$。当 $\theta=0$ 时，上式变为

$$V_k=\frac{1}{T^2W_k} \tag{4.148}$$

又因为

$$\langle m(t_0)\rangle=x(t)*v(t_0)$$

故

$$F[\langle m(t)\rangle]=T\cdot\underline{X}_k\cdot V_k=M_k$$

利用(4.148)式，则有：

$$M_k=\frac{\overline{X}_k}{TW_k} \tag{4.149}$$

在频率域上，根据(4.149)求出 M_k 之后，再对之作傅氏反变换即可得 $\langle m(t)\rangle$。

图 4.17 是一人工合成记录及其反演结果。此时，我们可以取 $M=256$，$\Delta=0.0125$。(a)是模型，(b)之右图是子波，左图是合成记录；

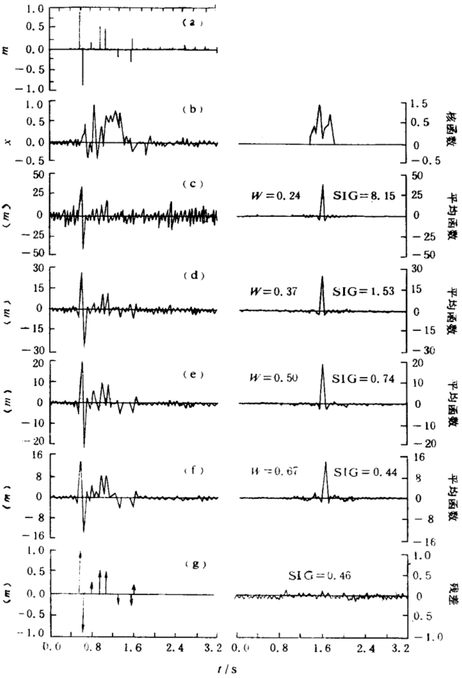


图 4.17　反褶积在频率域的实现

其(c)的左图是平均函数宽度 $w=0.24$，方差 SIG=8.15 时之反演结果；(d)，(e)，(f)，(g)都是加有 5%的白噪声的反演结果。从(d)到(g)，w 在不断加大，方差不断减小。可以看出，当方差和分辨力减小时，$\langle m(t)\rangle$渐渐接近于真实的反射系数 $m(t)$。

第五章　非线性反演问题

在前面几章中，我们系统地论述了解地球物理反演问题的线性反演法。这是一种理论最完整，应用最广泛而且效果也很理想的反演方法。然而，大多数地球物理问题都是非线性问题，因此研究非线性反演方法也是刻不容缓的任务。

随着地球物理学，特别是勘探地球物理学的发展，近年来，非线性反演法得到了迅速的发展。除了梯度法（Gradient method、the steepest descent、the steepest ascent）尝试法、（the trail and error method）、蒙特卡洛法（Monte Carlo method）等一些传统的非线性反演法外，许多新的反演方法，如人工神经网络（Artificial neural networks）法、模拟退火法（Simulated annealing）、遗传算法（Genetic Algorithm）、小波分析法（Wavelet analgsis）等等应运而生。随着计算技术的日新月异，特别是并行机的出现，需要大量计算时间的非线性反演法才如鱼得水，有了迅速发展的前提条件。

大家知道，所谓非线性问题，是指观测数据 $d_i(i=1,2,\cdots,M)$ 和模型参数 $m_j(j=1,2,\cdots,N)$ 之间不存在线性关系，这种非线性关系既可能呈显式 $\boldsymbol{d}=g(\boldsymbol{m})$，也可能呈隐式 $F(\boldsymbol{d},\boldsymbol{m})=0$。

解决这类非线性问题的方法，有如下两类：一类属于线性化的方法，即将非线性问题线性化，构成一种迭代的模式，用逐次逼近的方法求解，如二、三、四章所讲述的那些内容。另一类是不涉及非线性问题线性化，通过各种途径直接解非线性问题，实现从数据空间到模型空间的映射。

不管是那一类的反演问题，归根结底，反演过程都是一个对目标函数（或概率、概率密度）的最优化过程，只是实现最优的途径和方法

不同吧了。

本章中，我们将讨论目前在地球物理资料反演中常用的一些非线性反演方法。

§1 梯度法

梯度法(Gradient method)又称最速下降法(the steepest descent)或最速上升法(the steepest ascent)，是一种传统的非线性反演法，直到目前仍有许多地球物理资料的反演问题都采用梯度法求解。

在模型参数 $\boldsymbol{m}$ 和观测数据 $\boldsymbol{d}$ 呈隐式的情况下，有：

$$F(\boldsymbol{d},\boldsymbol{m})=0 \tag{5.1}$$

为方便起见，设

$$\boldsymbol{x}=[\boldsymbol{d}, \quad \boldsymbol{m}] \tag{5.2}$$

令

$$F(\boldsymbol{x})=F(\boldsymbol{d},\boldsymbol{m})=\begin{bmatrix} F_1(\boldsymbol{d},\boldsymbol{m}) \\ F_2(\boldsymbol{d},\boldsymbol{m}) \\ \vdots \quad \vdots \\ F_M(\boldsymbol{d},\boldsymbol{m}) \end{bmatrix}$$

如将这些非线性函数，构成如下目标函数：

$$\phi(\boldsymbol{x})=\sum_{i=1}^{M}[F_i(\boldsymbol{x})]^2 \tag{5.3}$$

显然，$\phi(\boldsymbol{x})$的零极小值点，就是方程(5.1)的解。

在多维空间中 $\phi(\boldsymbol{x})$函数构成一个高次曲面。为便于理解，现以二维空间为例，此时 $\phi(x_1,x_2)$所形成的曲面与 $x_1—x_2$ 平面相切的点就是它的零极小值点(图 5.1)。极小值点所对应的 x_1,x_2，就是观测数据 $\boldsymbol{d}$ 和模型 $\boldsymbol{m}$ 之对应值。

如果用 $\phi(\boldsymbol{x})=c_i(i=1,2,\cdots,k)$，这里 c_i 是常数，它相当于一系列平行于 $x_1—x_2$ 的平面。用这些平行平面切空间曲面 $\phi(\boldsymbol{x})=\phi(x_1,x_2)$，可以得到一族平面曲线，将它们投影到 $x_1—x_2$ 平面上，如图 5.2

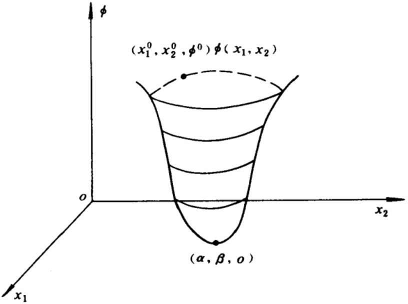


图 5.1　目标函数的示意图

所示，称为曲面的等高线族。且每一条线上的 ϕ 值均相同。由外向里，ϕ 值不断下降，当达到极小点时，ϕ 值为零。

在任意一个初始模型 $\boldsymbol{x}^{(0)}$ 处等高线的法线方向，就是 $\phi(\boldsymbol{x})$ 函数在该点处的梯度方向，即有：

$$\bar{\boldsymbol{g}}=\begin{bmatrix}\bar{g}_1\\ \bar{g}_2\\ \vdots\\ \bar{g}_P\end{bmatrix}=\frac{\partial\phi}{\partial\boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial\phi}{\partial x_1}\\ \dfrac{\partial\phi}{\partial x_2}\\ \vdots\\ \dfrac{\partial\phi}{\partial x_P}\end{bmatrix}$$

沿 $\bar{g}$ 的方向是 ϕ 值上升最快的方向。因此，其反方向为：

$$-g=-\frac{\partial\phi}{\partial x} \tag{5.4}$$

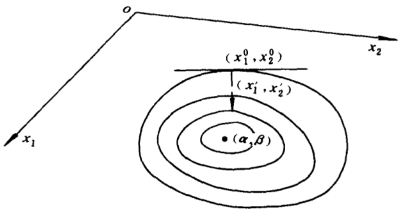


图 5.2 用等高线表示的目标函数

此式就是 ϕ 值下降最快的方向。梯度法，就是从一个初始模型出发，沿负梯度方向搜索求取 ϕ 函数极小点的一种最优化方法。

为了保证在校正过程中，ϕ 函数逐次下降，现选择

$$\boldsymbol{x}^{(i+1)}=\boldsymbol{x}^{(i)}-\lambda\overline{\boldsymbol{g}}^{(i)}=\boldsymbol{x}^{(i)}+\Delta\boldsymbol{x}^{(i)} \qquad (i=0,1,2,\cdots) \tag{5.5}$$

式中：$\boldsymbol{x}^{(i)}$ 表示第 i 次迭代之初始值；$\boldsymbol{x}^{(i+1)}$ 表示校正后的模型参数向量；$\Delta\boldsymbol{x}^{(i)}$ 是校正量，λ 是校正的步长因子。

当第 i 次迭代后，再从 $\boldsymbol{x}^{(i+1)}$ 出发，沿 $\phi(\boldsymbol{x}^{(i+1)})$ 的负梯度方向进行搜索，如此反复迭代，直至到达 $\phi(\boldsymbol{x})$ 函数的真正极小点为止。

步长因子 λ 如何选择？现作以下讨论。

将 $\phi(\boldsymbol{x}^{(i+1)})$ 按台劳级数展开，并略去高次项，即得

$$\begin{aligned}\phi(\boldsymbol{x}^{(i+1)})&=\phi(\boldsymbol{x}^{(i)})+\sum_{j=1}^{p}\frac{\partial\phi(\boldsymbol{x}^{(i)})}{\partial x_j}\Delta x_j^{(i)}\\&=\phi(\boldsymbol{x}^{(i)})+\overline{\boldsymbol{g}}^{(i)\mathrm{T}}\Delta\boldsymbol{x}^{(i)}\end{aligned} \tag{5.6}$$

式中：$p=M+N$ 是观测数据的个数 M 和未知参数的个数 N 之和。

将(5.5)式中的 $\Delta\boldsymbol{x}^{(i)}=-\lambda\overline{\boldsymbol{g}}^{(i)}$ 代入(5.6)式，则得：

$$\phi(\boldsymbol{x}^{(i+1)})=\phi(\boldsymbol{x}^{(i)})-\lambda\overline{\boldsymbol{g}}^{(i)\mathrm{T}}\overline{\boldsymbol{g}}^{(i)} \tag{5.7}$$

设 $\phi(\boldsymbol{x}^{(i+1)})=0$，也即经校正后使 ϕ 值达到了零极小值。则从(5.7)式

知：

$$\lambda=\frac{\phi(\boldsymbol{x}^{(i)})}{\boldsymbol{g}^{(i)\mathrm{T}}\boldsymbol{g}^{(i)}}=\frac{\phi(\boldsymbol{x}^{(i)})}{\sum_{j=1}^{p}\left(\frac{\partial\phi(\boldsymbol{x}^{(i)})}{\partial x_j}\right)^2} \tag{5.8}$$

按(5.8)式计算的 λ，代入(5.5)式后，如果

$$\phi(\boldsymbol{x}^{(i+1)})<\phi(\boldsymbol{x}^{(i)}) \tag{5.9}$$

成立，说明校正方向无误，则可继续进行迭代。否则，就减小 λ。例如每次减小一半，重复上述步骤，直至满足(5.9)式为止。然后，从 $\boldsymbol{x}^{(i+1)}$ 出发，重复上述过程，直至求得 ϕ 的零极小值为止。

一般说来，从任意初始模型出发，梯度法均能收敛。开始收敛速度快，往后越来越慢，尤其是在零极小值附近，要向极小点前进一步都必须付出较大的代价。因此，在实际应用过程中，常与牛顿法配合，两种方法相互取长补短，以达到既能保证收敛，又能加快迭代速度的目的。配合的方式多种多样，一般采用远离极小点时用梯度法，而当 ϕ 值降到一定程度后，改用牛顿法的方案。

一旦求出 $\phi(\boldsymbol{x})=0$ 的极小点，那么，相应的向量 $\boldsymbol{x}$ 就是其隐函数 $F(\boldsymbol{d},\boldsymbol{m})$ 的解。根据(5.2)式，不难求得模型参数 $\boldsymbol{m}$。

如观测数据 $\boldsymbol{d}$ 与模型 $\boldsymbol{m}$ 之间呈显函数时，可把目标函数设计为：

$$\phi=[\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{m})]^{\mathrm{T}}[\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{m})] \tag{5.10}$$

此后的做法和以上隐函数求解方法完全相同。然而，此时求出的只是待求模型 $\boldsymbol{m}$，而不是隐函数时的 $\boldsymbol{x}=(\boldsymbol{d},\boldsymbol{m})$。

下面，我们将从广义高斯分布的概率密度函数 $f(x)$ 的定义出发，进一步深入研究各种分布情况下的梯度法。

从概率论可知，具有 L_p 范数的概率密度函数为：

$$f_p(x)=\frac{p^{1-\frac{1}{p}}}{2\sigma_p\Gamma\left(\frac{1}{p}\right)}\exp\left[-\frac{1}{p}\frac{|x-x_0|^p}{(\sigma_p)^p}\right] \tag{5.11}$$

式中：$\Gamma\left(\frac{1}{p}\right)$表示伽马函数。

当 $p=1$ 时：

$$f_1(x)=\frac{1}{2\sigma_1}\exp\left[-\frac{|x-x_0|}{\sigma_1}\right], \tag{5.12}$$

可见，$f_1(x)$是中心在 $x=x_0$、平均偏差等于 σ_1 的对称指数分布。

当 $p=2$ 时：

$$f_2(x)=\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2}\frac{(x-x_0)^2}{{\sigma_2}^2}\right] \tag{5.13}$$

这里 $f_2(x)$是中心在 $x=x_0$ 处，标准偏差等于 σ_2 的高斯正态分布。

当 $p\rightarrow\infty$时，有

$$f_\infty(x)=\begin{cases}\frac{1}{2\sigma_\infty} & x_0-\sigma_\infty\leqslant x\leqslant x_0+\sigma_\infty \\ 0 & \text{其他 } x\end{cases} \tag{5.14}$$

可见，$f_\infty(x)$是中心在 $x=x_0$ 处，长为 $2\sigma_\infty$的箱状函数。

对同一事件 M 次独立观测数据而言，当 $p=1$ 时，

$$f_1(\boldsymbol{x})=\frac{1}{(2\sigma_1)^M}\exp\left[-\frac{1}{\sigma_1}\sum_{j=1}^{M}|x_j-x_0|\right] \tag{5.15}$$

与前几章所使用的符号一致，则上式变为：

$$f_1(\boldsymbol{m})=\frac{1}{(2\sigma_1)^M}\exp\left[-\frac{1}{\sigma_1}S_1(\boldsymbol{m})\right] \tag{5.16}$$

式中：

$$S_1(\boldsymbol{m})=\sum_{j=1}^{M}|d_j-g_j(\boldsymbol{m})| \tag{5.17}$$

显然，概率 $f(x)$最大，必然相应于 $S(\boldsymbol{m})$为最小。设：

$$\bar{g}_{ji}=\frac{\partial g_j(\boldsymbol{m})}{\partial \boldsymbol{m}_i}$$

则

$$r_{mi}=\frac{\partial S_1(\boldsymbol{m})}{\partial m_i}=\sum_{j=1}^{M}\bar{g}_{ji}\mathrm{sg}(d_j-g_j(\boldsymbol{m})) \tag{5.18}$$

式中：$\mathrm{sg}(x)$是符号函数，它具有：

$$\mathrm{sg}(x)=\begin{cases}1 & (x>0)\\ 0 & (x=0)\\ -1 & (x<0)\end{cases} \tag{5.19}$$

这里，由(5.18)式确定的就是目标函数 $S(\boldsymbol{m})$ 的正梯度方向。到此可知，就不难构成 L_1 范数条件下梯度法的反演框架。

在 M 个独立观测数据的情况下，当 $p=2$ 时，

$$f_2(\boldsymbol{m})=\frac{1}{(2\pi)^{M/2}\sigma_2{}^M}\exp\left[-\frac{\sigma_2^{-2}}{2}\sum_{j=1}^{M}(d_j-g_j(\boldsymbol{m}))^2\right] \tag{5.20}$$

和 L_1 范数情况相似，只有目标函数

$$S_2(\boldsymbol{m})=\sum_{j=1}^{M}(d_j-g_j(\boldsymbol{m}))^2$$

极小，才能使 $f_2(\boldsymbol{m})$ 极大。前面已经论述 L_2 范数情况下梯度法的反演原理，故不再赘述。

对任意 p 而言，由(5.11)式可知，$f_p(\boldsymbol{m})$ 最大，在 M 次独立观测的情况下，必然是函数

$$R_p(\boldsymbol{m})=\frac{1}{p}\sum_{j=1}^{M}\frac{|d_j-g_j(\boldsymbol{m})|^p}{\sigma_p{}^p} \tag{5.21}$$

为最小，或

$$S_p(\boldsymbol{m})=pR(\boldsymbol{m})=\sum_{j=1}^{M}\frac{|d_j-g_j(\boldsymbol{m})|^p}{\sigma_p{}^p} \tag{5.22}$$

为最小。当 $p\to\infty$ 时，

$$S_\infty(\boldsymbol{m})=\max\left[\frac{d_j-g_j(\boldsymbol{m})}{\sigma_{\infty j}},(j\in I_M)\right] \tag{5.23}$$

当 $j=k$ 时，$S_\infty(\boldsymbol{m})$ 达到极大，则有：

$$\bar{g}_{ki}=\frac{\partial g_k(\boldsymbol{m})}{\partial m_i}$$

所以：

$$r_{m_i}=\frac{\partial S_\infty(\boldsymbol{m})}{\partial m_i}=\bar{g}_{ki}\mathrm{sg}(d_k-g_k(\boldsymbol{m})) \tag{5.24}$$

就是函数 $S(\boldsymbol{m})$ 的正梯度方向。由此，不难得到 L_∞ 时梯度法之反演

步骤。

简而言之，对 L_p 范数而言，梯度法的参数修改方式如下：

$$m_j^{(i+1)}=m_j^{(i)}-r_{m_j}^{(i)}\Delta m_j^{(i)} \qquad (j=1,2,\cdots,N) \tag{5.25}$$

式中：

$m_j^{(i+1)}$，$m_j^{(i)}$ 分别代表第 i 个参数 m_j 在第$(i+1)$次迭代（修改后）和第 i 次迭代（修改前）之值；$r_{m_j}^{(i)}$ 和 $\Delta m_j^{(i)}$ 为 m_j 在第 i 次迭代中之梯度方向以及修改量。

§2 尝试法

设

$$\boldsymbol{d}=\boldsymbol{g}(\boldsymbol{m})$$

所谓尝试法，就是解释人员从一个初始模型 $\boldsymbol{m}_0$ 出发，计算 $\boldsymbol{g}(\boldsymbol{m}_0)$，并且把它与实际观测值 $\boldsymbol{d}_{obs}$比较。然后，解释人员从其已知的物理知识和先验信息出发，猜测另一个新的模型 $\boldsymbol{m}_1$ 以使它拟合观测数据 $\boldsymbol{d}_{obs}$比 $\boldsymbol{m}_0$ 拟合要好。重复上述步骤，直到后续模型之理论值与观测值的方差不再减小或减小拟合误差十分缓慢（符合要求）为止。

用尝试法时，拟合好坏可以有不同定义，如用基于 $L_1,L_2,\cdots,L_\infty$ 范数的目标函数等，当然拟合好坏也可以靠定性估计。

通常，尝试法可在计算机终端上通过人机对话来实现。在迭代过程中，解释人员可以从屏幕上直接观察到曲线拟合好坏，了解到参数是否符合已知的先验信息。

尝试法的主要优点是：

(1)不需解决更多的数字问题；

(2)不需要较大的内存；

(3)可以设置衡量观测数据拟合好坏不同的目标函数；

(4)由于采用人机对话，给人以十分直观的感觉。

不足之处是：对参数较多的反演问题计算时间长，迭代次数多，

且难以操作。

简言之，尝试法只对那些对反演问题的解有先验知识，且计算机能力较强的解释人员，对待求模型的参数不多的情况才适用。

§3 蒙特卡洛法

如前所述，多数非线性反演方法都需要充分地暴露模型空间，进行大量计算，才可能找到比较合理的解。但由于计算工作量太大，因而往往难以使用。

如果问题的非线性并非十分严重，模型参数较少，目标函数仅有单个极小，这类反演问题可用梯度法等求解。对于高次非线性函数，模型参数较多，且目标函数具多个极小的情况而言，用诸如梯度法等简单非线性反演方法，有着巨大的危险性。这时，在反演的过程中容易陷入局部极小。有戏剧性的是，如果我们随机在模型空间中选择模型，以求得总体极小，比规则地划分模型空间，以求出总体极小，所需的计算时间和耗费的经费都要少。

为纪念有名的赌城——蒙特卡洛，人们将反演过程中任何一个阶段，用随机（或似随机）发生器产生模型的方法统称为蒙特卡洛法。因此，蒙特卡洛法实际上就是一种"赌博法"。它可以用来解决高次非线性的、多参数、具有多个局部极小的非线性反演问题。

假定，解释人员对待求模型参数的范围有一些先验信息，即：

$$m^{\alpha}_{\min} \leqslant m^{\alpha} \leqslant m^{\alpha}_{\max} \qquad (\alpha \in I) \tag{5.26}$$

式中：α 代表参数的个数；$m^{\alpha}_{\min}$和 $m^{\alpha}_{\max}$分别代表模型参数 m^{α} 之下界和上界。

蒙特卡洛法的步骤如下：

(1)用随机发生器在(5.26)式范围内产生一个模型 $\boldsymbol{m}$；

(2)计算 $\boldsymbol{d}=\boldsymbol{g}(\boldsymbol{m})$；

(3)比较理论值 $\boldsymbol{d}$ 和观测值 $\boldsymbol{d}_{\text{obs}}$，并利用一定准则判断 $\boldsymbol{m}$ 是否可以接受；

(4)如果 $\boldsymbol{m}$ 可以接受，则停止计算，否则重复(1)～(3)步骤，直至满足要求为止。

例如，Press 在 1968 年发表的文章中，利用蒙特卡洛法研究了地球地幔的密度以及纵、横波的速度。他利用的资料是地球振动的本征周期，地震波在旅行时的地球的质量以及转动惯量等。此时，待求的参数是密度 $\rho(r)$，纵波速度 $\alpha(r)$ 和横波速度 $\beta(r)$。反演时，将地球半径划分为 23 分(其余由内插法确定)。因此，参数总数是 $23\times3=69$ 个。即

$$\begin{aligned}\boldsymbol{m} &= (m^1, m^2, \cdots, m^{69}) \\ &= (\rho(r_1), \cdots, \rho(r_{23}), \alpha(r_1), \cdots, \alpha(r_{23}), \beta(r_1), \cdots, \beta(r_{23}))\end{aligned}$$

在模型空间中，近似有 500 万个模型进行了试算，其中有 6 个可以接受的模型。图 5.3 是其中的三个模型，计算的结果与我们已知的地球模型十分相近。

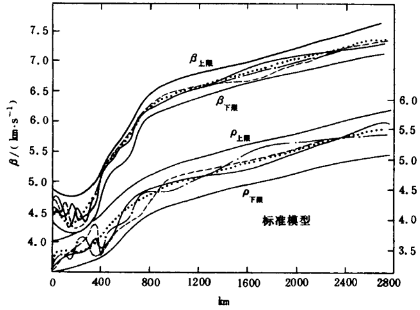


图 5.3　蒙特卡洛法求出的地球的速度、密度模型

§4 人工神经网络(ANN)法

人工神经网络，又称神经计算机与并行机，近年来已成国内外的一个热门研究课题。在地球物理资料处理和反演解释中得到卓有成效的应用。

由于现行计算机只能解决那些有严格数学定义的形式思维和逻辑思维问题，而对人脑许多更高级、更大量、更复杂的形象思维、灵感思维和创造性思维就无能为力了。

为了让读者了解神经网络的基本原理及其在地球物理资料反演中的应用，我们先从神经网络的基本特征讲起。

1. 神经网络的基本特征

如果将人脑神经信息活动的特点与现行计算机工作方式作一比较，可以看出：

(1) 人在识别一幅图像或作出一项判断决策时，并不像现行的计算机那样，按预先编好的程序一条接一条地执行指令，而是运用存在于脑中多方面的知识和经验同时并发以迅速作出解答。巨量并行性是人脑信息活动的一个重要特征。

(2) 计算机存储地址和存储的内容是彼此分开的，而人脑中信息存储和信息处理是合在一起的，这是人脑信息处理的又一重要特征。

(3) 人脑不像计算机那样，只能被动地执行已编好的程序，而是能够通过内部自组织、自学习的能力，不断地适应外界环境，有效地处理各种模拟的、模糊的或随机的问题。人脑这种自组织、自学习的功能，是现代电脑所望尘莫及的。

目前已清楚的适宜利用神经网络和神经计算机的领域有：模式识别、信号处理、判断决策、组合优化，知识工程等。在地球物理数据的处理和反演中也得到了成功的应用。

2. 简单人工神经元模型

由于现代科学技术的发展，人们对神经的理解已有了相当深度。由于实验手段的局限性和人脑的特殊性，建立大脑模型的研究方法是行之有效的方法之一。这里，我们将讨论几个简单的人工神经元模型。

(1)M-P 神经元模型。M-P 模型是 Mcculloch-Pitts 1943 年提出来的，它是一个多输入单输出的非线性系统(见图 5.4)。

设某一神经元的输入为：

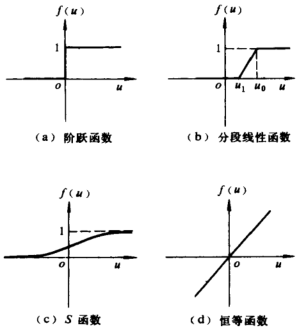


图 5.4 M-P 神经元模型

$$\boldsymbol{x}=[x_1,x_2,\cdots,x_n] \tag{5.27}$$

它们相应的权值为：

$$\boldsymbol{w}=[w_1,w_2,\cdots,w_n] \tag{5.28}$$

如神经元的阈值为 θ，则输出 y 可表示为：

$$y=1\left(\sum_{j=1}^{n}w_jx_j-\theta\right)=1(\boldsymbol{wx}-\boldsymbol{\theta}) \tag{5.29}$$

式中"1"表示单位函数，即

$$1(u)=\begin{cases}1 & (u\geqslant 0)\\ 0 & (u<0)\end{cases} \tag{5.30}$$

可见，M-P 模型的特点是：

①多输入，单输出；

②阈值作用；

③输入与输出均为两态(抑制、兴奋)；

④每个输入通过权值来表征它对神经元之耦合程度(如无耦合可取 $w_j=0$)。

(2)连续神经元模型(见图 5.5)。为反映神经元状态参数连续变化的情况，常用一阶非线性微分方程来模拟生物神经元膜电位随时间变化的规律，即

$$\begin{cases}\tau\dfrac{\mathrm{d}u(t)}{\mathrm{d}t}=-u(t)+\displaystyle\sum_{j=1}^{n}w_jx_j(t)-\theta\\ y(t)=f(u(t))\end{cases} \tag{5.31}$$

式中：τ 为时间常数；θ 为静止膜电位；$f(u)$ 为输入输出函数。它有四种可能形式，即

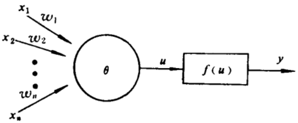


图 5.5　连续神经元模型

① $f(u)=\begin{cases}1 & u\geqslant 0\\ 0 & 0<0\end{cases}$

② $f(u)=\begin{cases}1 & u\geqslant u_0\\ au+b & u_1\leqslant u<u_0\\ 0 & u<u_1\end{cases}$

③ $f(u)=\dfrac{1}{1+\exp(-u+c)}$

④ $f(u)=u$

3. 神经网络的计算机理

神经网络计算机的计算机理与电脑不同。其主要特征如下：

(1) 处理信息的高度并行性。现行计算机的信号输入与数据处理都是串行的，一个一个信号输入，输入之后，又按已编程序的解题方法，一个指令接着另一个指令，只有当最后一条指令完成后，才得到最后的结论。而人工神经网络或人脑却不同，信号的输入和处理都是并行的，相当于同时输入，同时处理。因而，它能以快得多的计算方式作出判断或求得解答。

(2) 信息处理和信息存储合二为一。与电脑不同，神经元既是处理单元，又是信息存储单元。信息处理的结果反映在突触(synaptic)——神经元之间的连接键——连接强度的变化上。"突触"既是转换站，又是信息的存储站。所以计算时，它不像电脑那样先找存储地址，后提取存储内容。

(3)能接受和处理模拟的、模糊的和随机的信息。

(4)求满意解而不是精确解。因为人类日常生活中，许多都不是精确解，而是满意解。

(5)具有自组织和自学习的能力。

在简单介绍人工神经元一般知识后，下面我们将就在地球物理资料反演中常用的 Hopfield 网络模型及其应用加以介绍。

4. Hopfield 网络及其在地球物理资料反演中的应用

1985 年，霍普菲尔德(J. J. Hopfield)和塔克(D. W. Tank)建立

了互相连结型神经网络模型。在 Hopfield 网络中，每一个神经元都和其他神经元相连接，所以又称为全互连接网，神经元之间的连接权 w_{ij}满足

$$w_{ii}=0 \qquad (i=1,2,\cdots,N) \tag{5.32}$$

$$w_{ij}=w_{ji} \qquad (i,j=1,2,\cdots,N) \tag{5.33}$$

式中：N 是神经元的个数。

在离散 Hopfield 网络中，如图 5.6 所示，$v_1,v_2,\cdots,v_N$ 为神经元 i 的输入，它们对第 i 个神经元的影响程度用连接权 $w_{i1},w_{i2},\cdots,w_{iN}$来表征；$\theta_i$ 为神经元 i 的阈值；v_i 为其输出，则有：

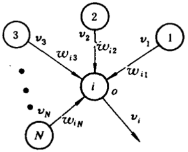


图 5.6 离散型 Hopfield 网络中神经元工作原理图

$$v_i=\mathrm{sgn}\left(\sum_{\substack{j=1\\i\neq j}}^{N} w_{ji}v_j-\theta_i\right) \tag{5.34}$$

显然，Hopfield 网络是一个多输入、多输出、带阈值的二态非线性动力系统。且

$$v_i=\begin{cases}1\\0\end{cases} \qquad D_i=\sum_{\substack{j=1\\i\neq j}}^{N} w_{ji}v_j-\theta_j=\begin{cases}>0\\\leqslant 0\end{cases} \tag{5.35}$$

定义 Hopfield 网络的能量函数为：

$$E = -\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\i\neq j}}^{N} w_{ij}v_iv_j + \sum_{i=1}^{N}\theta_iv_i \tag{5.36}$$

则其变化量为：

$$\Delta E_i = -\frac{\partial E}{\partial v_i}\Delta v_i = \Delta v_i\left(-\sum_{\substack{j=1\\i\neq j}}^{N} w_{ij}v_j + \theta_i\right) \tag{5.37}$$

如设 $\Delta v_i = v_i(t) - v_i(t-1)$。这里 $v_i(t-1)$ 为前一时刻之输出，$v_i(t)$ 为此时之输出。由(5.34)式不难看出 v_i 只能为 0 或 1。若 $\Delta v_i < 0$，就意味着 $v_i(t-1)=1$，而 $v_i(t)=0$，由(5.35)式可知，此时 $D_i \leqslant 0$。因此，必然有 $\Delta E_i = \Delta v_i(-D_i) \leqslant 0$；若设 $\Delta v_i > 0$，就意味着 $v_i(t-1)=0$，而 $v_i(t)=1$。由(5.35)式可知，此时 $D_i > 0$ 根据(5.37)式必然有 $\Delta E_i < 0$。结论是，在 Hopfield 网络迭代过程中，总有 $\Delta E_i \leqslant 0$，网络总收敛到能量极小的状态。换言之，计算能量总是不断地随第 i 个神经元状态的变化而下降。同理，可对其他神经元的状态变化作类似的分析。

Hopfield 网络的最重要的应用之一是最优化问题。这种应用的关键在于：将 Hopfield 网络的能量函数和最优化问题（例如，地球物理反演问题）的目标函数联系起来，找到地球物理反演问题中的模型 $(m_i, i=1,2,\cdots,M)$、核函数 $(G_{ij}, i,j=1,2,\cdots,M)$ 和神经元诸要素，如神经元的输入、输出和它们之间的连结权等，在神经元稳定输出状态下之关系。

Hopfield 网络在地球物理反演中应用的另一关键问题是，如何将地球物理的模型参数用二进制的形式表示。地球物理问题不同，这种表达形式也不尽相同。

下面，我们将以前面几章中讲述的线性反演问题为例，对上述两个问题加以说明。

设有目标函数：

$$\psi = \frac{1}{2}\sum_{k=1}^{M}\left(d_k - \sum_{i=1}^{N} G_{ki}m_i\right)^2$$

$$=-\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\i\neq j}}^{N}\left[-\sum_{k=1}^{M}G_{kj}G_{ki}\right]m_i m_j$$

$$-\sum_{i=1}^{N}\left[-\frac{1}{2}\sum_{k=1}^{M}G_{ki}^2 m_i+\sum_{k=1}^{M}G_{ki}d_k\right]m_i+\frac{1}{2}\sum_{k=1}^{M}d_k^2 \qquad (5.38)$$

和(5.36)比较，就可发现，除常数项$\frac{1}{2}\sum_{k=1}^{M}d_k^2$外，如设：

$$v_i=m_i$$

$$w_{ij}=\left[-\sum_{k=1}^{M}G_{ki}G_{ij}\right] \qquad (5.39)$$

$$\theta=-\left[-\frac{1}{2}\sum_{k=1}^{M}G_{ki}^2 m_i+\sum_{k=1}^{M}G_{ki}d_k\right] \qquad (i=1,2,\cdots,N)$$

则二式完全相似。如果模型$m_i(i=1,2,\cdots,N)$是等于0或1的序列，不难看出，在Hopfield网络中，只需N个神经元就可以了。如果，模型参数是任意实数，为了用Hopfield网络进行反演，还必须将模型参数的表示形式加以改变，如下式所示的二进制表达式，即

$$m_i=\sum_{j=-D}^{U}2^j B_{ij}-2^U B, \qquad (i=1,2,\cdots,N) \qquad (5.40)$$

式中：$B_{ij}=0$或1；D和U是分别取决于模型参数的精度和大小的整数；而B数值上等于1，代表B_{ij}的维数。方程(5.40)式，可以看成模型$m_i(i=1,2,\cdots,N)$的二进制表达式。

将(5.40)式代入，(5.38)式，则得：

$$\psi=-\frac{1}{2}\sum_{i=1}^{N}\sum_{m=-D}^{U}\sum_{j=1}^{N}\sum_{n=-D}^{U}\left[-\sum_{k=1}^{M}2^{(m+n)}G_{ki}G_{kj}\right]B_{im}B_{jn}$$

$$-\sum_{i=1}^{N}\sum_{m=-D}^{U}\left[-\frac{1}{2}\sum_{k=1}^{M}(2^m G_{ki})^2 B_{im}+\sum_{k=1}^{M}(2^m G_{ki})d_k\right.$$

$$\left.+\sum_{k=1}^{M}\sum_{j=1}^{N}(2^{(m+u)}\cdot G_{ki}G_{kj})B\right]\cdot B_{im}+\alpha \qquad (5.41)$$

式中：α是与模型无关的一常数项。

比较(5.41)式和(5.36)式可以看出，除 α 以外，两式具有相同的形式，如果设

$$v_{im}=B_{im} \tag{5.42}$$

$$T_{im,jn}=\left[-\sum_{k=1}^{M}2^{m+n}G_{ki}G_{kj}\right] \tag{5.43}$$

且 $T_{im,jn}=0$ 当 $(i,m)=(j,n)$

$$\theta_{im}=-\left[-\frac{1}{2}\sum_{k=1}^{M}(2^{m}G_{ki}^{2})B_{im}+\sum_{k=1}^{M}(2^{m}G_{ki})d_{k}\right.$$
$$\left.+\sum_{k=1}^{M}\sum_{j=1}^{N}(2^{(m+n)}\cdot G_{ki}G_{kj})B\right] \tag{5.44}$$

在这种条件下，(5.36)式变为：

$$E=-\frac{1}{2}\sum_{i=1}^{N}\sum_{m=-D}^{U}\sum_{j=1}^{N}\sum_{n=-D}^{U}T_{im,jn}v_{im}v_{jn}+\sum_{i=1}^{N}\sum_{m=-D}^{U}v_{im}\theta_{im} \tag{5.45}$$

显然，网络中神经元的数目等于 $N(U+D)$，而 D 和 U 是充分大的正整数，以保证能准确地表示模型参数的幅值和精度。

到此为止，我们已详细地论述了在地球物理反问题中应用 Hopfield 网络所遇到的两个关键问题。下面，让我们分析一下，利用 Hopfield 网络进行线性反演的主要步骤。

第一步，根据已知地质、地球物理信息，确定地球物理模型参数(如厚度、速度、密度、电阻率等)之可能变化范围，并根据初始模型参数，确定表征它们的 $U,D,B_{ij}(i=1,2,\cdots,N;\quad j=1,2,\cdots,D+U)$ 和 B。

第二步，根据初始模计算的核函数 $G_{ij}(i=1,2,\cdots,M;\quad j=1,2,\cdots,N)$ 和上一步骤中确定的 $U、D、B_{ij},B$，按(5.42)式～(5.44)式计算第 i 个模型参数相应的 $(U+D)$ 个神经元之输入 $v_{in}(i=1,2,\cdots,N,\quad n=1,2,\cdots,(U+D))$，神经元之间的连结权 $T_{im,jn}(i,j=1,2,\cdots,N;\quad m,n=1,2,\cdots,(U+D))$，以及阀值 $\theta_{im}(i=1,2,\cdots,N;\quad m=1,2,\cdots,(U+D))$ 等进行 Hopfield 网络运算。计算每个模型参数 $(m_i,i=1,2,\cdots,N)$ 对应的 $(U+D)$ 个神经元之输出 $v_{im}(i=1,2,\cdots,$

N； $m=1,2,\cdots,U+D$)。

第三步，根据计算的输出 v_{im}，按(5.42)式求出 B_{im}，再由已知的 U,D,B 计算 i 个模型参数之值。

第四步，按(5.41)计算拟合方差 ψ 的值，如 ψ 值小于给定的某一正数 ε，则将第三步求得的模型参数作为反演结果，否则重复第一到第三步运算，直至符合要求为止。

必须指出，为了适应神经元的要求，模型参数 $m_i(i=1,2,\cdots,N)$ 的表示式应作相应的变化，如(5.40)式所示的二进制表达式(但(5.40)只是一种表达形式，模型参数不同，其幅值和变化范围各异，表达式就应该有所变化。

用 Hopfield 网络进行地球物理资料反演时，也同样存在收敛到局部极小的可能性，初始模型的选择对反演结果关系很大，应该尽量初始模型之逼近待求的真实模型，这样既可以减少计算时间，又可避免陷入局部极小。

除 Hopfield 网络外，有些学者也曾把 BP 网络利用于地球物理资料的反演，特别是非线性反演，读者若有兴趣可参考有关文章。

§5 模拟退火法(Simulated Annealing 或 SA)

物体加热、熔化，能量增加；降温、冷却，即退火时，能量减少。如果冷却十分缓慢，物体就会形成理想的晶体，其能量总体出现极小；反之冷却并非十分缓慢，就会出现物体能量的局部极小，成为似稳状的玻璃状物质。退火过程类似于利用最优化原理解地球物理反演问题。在退火过程中，欲求能量最小的理想晶体，并非要形成能量局部极小状态下的玻璃状物，而反演的目的是求取目标函数能量总体极小状态下的解，而不是收敛到局部极小状态下模型的解。基于这种类似，Rothman(1985,1986)将退火原理引入地球物理资料的反演，并称之为模拟退火法。

据统计热力学，物体中每个分子的状态服从吉布斯概率分布，即

为：

$$\rho(r_i)=\frac{\exp\left(-\frac{E(r_i)}{k_bT}\right)}{\sum_{j=1}^{M}\exp\left(-\frac{E(r_j)}{k_bT}\right)} \tag{5.46}$$

式中：$E(r_i)$为第 i 个分子的能量函数；r_i 为第 i 个分子所处的状态；$\rho(r_i)$为其概率密度；k_b 为玻尔兹曼常数；T 为温度。

从上式可以看出，随着温度的降低，分子的能量随之变小，物体从融熔的液态向固态转变，每个分子取哪种状态由其概率密度的大小来决定。概率密度越大，分子取该状态的可能性就越大。

模拟退火法用于反演的基本思路是将待反演的模型参数看作是熔化物体的每一个分子，将目标函数看作是熔化物体的能量函数，通过缓慢地减小模拟温度来进行迭代反演，使目标函数最终达到极小。

模拟退火法有两种算法，即 Metropolis 算法（简称 MSA）和 Heat bath 算法（简称 HBSA）。两者的区别在于搜索模型空间和模型参数修改量不同。MSA 算法可在全空间自动搜索，模型修改量是随机的；而 HBSA 算法则是把模型参数限制在一定范围内，模型修改量是某一固定值，两种算法本质上是一致的。研究表明，HBSA 的计算速度比 MSA 快，特别是在已知某些模型参数的情况下，由于模型空间的缩小，计算速度就进一步加快了。

在 MSA 算法中，若待定模型参数 $m_i(i=1,2,\cdots,N)$，反演时先取一初始模型 $m_i^l(i=1,2,\cdots,N)$，计算相应的目标函数，即

$$E(\boldsymbol{m}^l)=\sum_{j=1}^{M}(d_j-f(\boldsymbol{m}^l))^2$$

然后，随机地修改模型，得 $\boldsymbol{m}^{l+1}$，其对应的目标函数为 $E(\boldsymbol{m}^{l+1})$。并设

$$\Delta E=E(\boldsymbol{m}^{l+1})-E(\boldsymbol{m}^l)$$

则由以下方法判断模型参数修改量是否可以接受，即

$$\rho(\Delta E)=\exp\left(-\frac{\Delta E}{k_bT}\right) \tag{5.47}$$

若 $\Delta E \leqslant 0$，表明模型修改方向使目标函数减小，修改可以接受；若 $\Delta E > 0$，按(5.46)式计算 $\rho(\Delta E)$；，若 $0 < \rho(\Delta E) < R$，说明修改仍可接受，否则拒绝修改，这里 R 是在 0 和 1 之间的一个随机数。

从(5.47)可以看出，温度 T 只是一个控制参数，玻尔兹曼常数 k_b 只起一个尺度因子的作用，实际应用时取为 1。显然，T 较小时，对 ΔE 起着一种放大作用，使低温时不易接受模型修改。这相当于物体在低温时分子被束缚在平衡位置附近。

HBSA 算法也是随机地选择一个初始模型 $\boldsymbol{m}^l$，其参数分别为 $m_1^l, m_2^l, \cdots, m_N^l$。在保持 $m_2^l, \cdots, m_N^l$ 固定的情况下，使 m_1^l 从 $m_1^{l\min} \sim m_1^{l\max}$ 取出所有可能的取值，按(5.46)计算概率密度，并选择概率最大之 m_1^l 为其新值，记为 m_1^{l+1}。接着又固定 $m_1^{l+1}, m_2^l, \cdots, m_N^l$，按 m_1 的方法选择 m_2^l 的新值……选择概率最大者，并记为 $m_2^{(l+1)}$。依此类推，把完成 $m_1^l \sim m_N^l$ 的一次选择称为一次迭代。在每一次迭代过程中保持温度 T 不变，温度 T 随迭代次数的增加呈指数降低。随着温度 T 的降低，$m_1 \sim m_N$ 最终达到一个稳定解。

显然，在模拟退火法中，退火温度的选取最为重要，选择不当，将导致反演失败。根据前人研究，T 取指数变化的模式，比较切合退火的实质，即

$$T(k) = T_0 (0.99)^k \tag{5.48}$$

式中：k 为迭代次数；T_0 为初始温度（T_0 应视具体情况而定）；$T(k)$ 为第 k 次迭代时之温度。

用模拟退火法反演地球物理资料仍然存在收敛于局部极小的问题，这有待进一步深入研究。与其他非线性反演相比，模似退火法不依赖于初始模型的选择，同时在反演过程中不需要计算雅可比矩阵。因此，在各种地球物理资料的反演中都有许多成功的实例。

§6 遗传算法(Genetic Algorithm 或 GA)

遗传算法也称基因算法，是模拟自然选择和遗传学理论，依据适

者生存的原理而建立的一种最优化方法。

自然界中生物进化的动力是遗传和自然选择。遗传是通过基因的复制和基因的突变来进行的。基因附着在染色体上，染色体总是成对出现，交配时成对的染色体分裂，父母双方各复制自己的每对染色体中的一个传给下一代，这就是基因的复制。复制使得父代的特性能在子代上体现并保持下去。子代的基因有可能发生突变，即一种基因突然转变成另一种类型的基因。基因突变引起生物体的突然变化，这对生物的进化影响很大。生物繁殖的下一代是否能继续繁衍下去，是受自然选择规律支配的，即物种间的相互竞争、表现出优胜劣汰。

遗传算法就是模拟上述生物进化过程。它首先通过随机选择生存一个模型群体（相当于生物中的种群），这个模型集有 N 个成员，把它们作为初始模型；然后用生物进化的方法，使初始模型最终进化到全局最优解。具体做法作如下介绍。

（1）二进制编码形成染色体。遗传算法不直接处理模型参数，而是对模型参数的二进制编码进行操作。将待反演的每一个参数进行二进制编码（即将十进制的数转化为二进制的数），将此二进制码称为“染色体”，多个模型参数对应多条染色体。每一个二进制码代表一个“基因”，它只能取 0 或 1。

一个模型参数都有一定的变化范围，一个简单而又实用的方法是把模型参数的可能空间按等差方式编码，即有：

$$m_{ij}=\min(m_j)+i\Delta m_j \qquad (5.49)$$

式中：$\min(m_j)$表示第 i 个模型参数的最小值；Δm_j 为第 i 个模型参数 m_j 的分辨力；m_{ij}表示第 i 个模型 m_j 的第 i 个取值。该取值对应着一个二进制码（即一个染色体）。显然，二进制码的长度由模型参数的范围和分辨力决定，如图 5.7(a)

（2）配对成员的选择。根据初始模型集中各成员的拟合函数值（目标函数值）的大小，以一定的选择概率，使初始模型集中的所有模型都相互配对，配对的两个模型将作为父代模型来繁殖下一代。

(3) 基因交换。模拟生物进化过程的基因复制，即将配对模型相对应的染色体，以一定交换概率随机地分裂，并互相交换为图 5.7(b)所示，使基因进行重组，形成新的一代。若每次只对一个染色体进行交换，称"单点交换"，对多条染色体进行交换，则称"多点交换"(一般采用后者)。从生物学角度看，两个配对模型可以生成许多个子代模型。但是，用计算机模拟时，为防止数量成指数增加而使计算机无法存储，一般说来，在实现时，两个配对模型只生存两个子代模型。

(a) 二进制模型编码

*	*	*	*	*	*	*

$m_{ij}=$	0	0	0	0	0	0	0	$\min_{ij}$
$m_{ij}=$	0	0	0	0	0	0	1	$\min_{ij}+1\Delta m_{ij}$
$m_{ij}=$	0	0	0	0	0	1	0	$\min_{ij}+2\Delta m_{ij}$
$m_{ij}=$	0	0	0	0	0	1	1	$\min_{ij}+3\Delta m_{ij}$
	…	…						…
$m_{ij}=$	1	1	1	1	1	1	1	$\max_{ij}$

m_{ij}=第 j 次迭代第 i 个模型参数

$\min_{ij}$=第 j 次迭代第 i 个模型参数的最小值

$\max_{ij}$=第 j 次迭代第 i 个模型参数的最大值

Δm_{ij}=第 j 次迭代第 i 个模型参数的分辨率

(b) 基因交换

*	*	*	*	1	0	0	m_{ij}
*	*	*	*	0	1	1	m_{ij}^{*}

(c) 基因变异

*	*	1	*	*	*	*	m_{ij}
*	*	0	*	*	*	*	m_{ij}^{*}

图 5.7 遗传算法示意图

(4) 基因突变。模拟生物进化过程中的基因突变，以一定的变异

概率，在基因交换操作形成的两个子代模型中随机地使其中一个的某条染色体的基因发生突变，即使原来为 1 的基因变为 0，或原来为 0 的基因变成 1（如图 5.7(c)所示）。研究表明，较小的变异概率在反演过程中十分必要，它能产生某些有效的基因片段。

(5)模型的更新。模拟生物进化过程的自然选择，经过基因交换和变异，两个旧的模型生存了两个新的子代模型。新的模型继承了旧模型中的某些特征。如把整个旧模型都用新模型代替，则称为分代型遗传算法；如果只替代拟合程度较差的旧模型，则称为稳态型遗传算法。前者收敛速度慢，后者收敛速度快，故一般采用后者。

上述的基因交换、变异和模型的更新是对初始模型集中的每一对配对模型进行操作，待所配对模型操作完后，初始模型集便进化到第一代模型集（相当于一次迭代）。将第一代模型集当作初始模型集重复上述步骤，便可以得到第二代模型集，如此往复以致第三代、第四代模型集…。上述迭代过程的终止条件是直至模型集中的每一个成员的拟合函数值基本相等，也即拟合函数值的平均值趋于群体的最大拟合函数值。此时，对模型参数进行二进制解码即为反演结果。

研究表明，遗传算法并不能绝对保证目标函数收敛到全局最优解。但只要模型群体的成员数的大小以及交换和变异概率选择合适，这种算法一般不会陷入局部极小。据文献介绍，较大的更新概率(0.9)、中等程度的交换概率(0.6)和较小的变异概率(0.01)相结合时，遗传算法的效果最佳。

和其他反演算法相同，遗传算法也是从一初始模型集出发，计算理论曲线与实测曲线的拟合情况，并根据拟合结果通过遗传操作对模型参数进行各种各样的修改，使模型集向最优解逐步演化。其遗传的代数相当于迭代的次数。在迭代过程中对配对模型的染色体进行交换和变异操作，实际上就是对模型空间进行多途径的搜索。由于是对每一条染色体（即一个模型参数）同时进行操作，所以算法本身具有平行性，搜索效率比一般非线性反演方法高。

§7 多尺度反演(Multi-Scale Inversion 或 MSI)

多尺度反演是近年来发展起来的一种基于小波分析中的多尺度分解理论的反演方法,受到地球物理学界的广泛关注。

多尺度反演是一种加快收敛速度,克服局部极值影响,搜索全局最小极值点的一种反演手段。其目的之一是为了克服许多非线性反演常受众多局部极小之困扰。

多尺度反演是把目标函数分解成不同尺度的分量,根据不同尺度上目标函数的特征逐步搜索全局最小点。一般情况下,在大尺度(或低波数)上,目标函数极值点少,且分得很开。用通常的方法很容易直接或经比较搜索出大尺度(总体背景)上的“全局极小点”。在相对较小尺度(稍大波数)上,目标函数极值点较多,直接寻找全局极值点,虽比作尺度分解之前容易,但仍然比较困难。不过,如果我们以大尺度搜索到的背景“全局极小点”为起始点,在其附近继续搜索,也能容易地搜索到中等尺度上的“全局极小点”。如此,不断地缩小尺度,在不断加入高分辨成分的目标函数上,调整全局最小点,最终当目标函数的尺度降至原始尺度时,对应搜索出的全局最小点,就是真正的全局最小点。这种做法的优点是,在大尺度(或低波数)上,反演稳定,反演结果不受初始模型的影响,能避免其后的反演落入错误的领域,使收敛速度加快。

1. 尺度的概念

尺度是指当我们以离散方式描述某一空间(或时间)函数时,均匀离散点之间的距离。尺度是分辨率的倒数。分辨率被定义为单位距离内离散点的个数。因此,若分辨率为 $2^{-j}(j\in Z)$则对应的尺度为 2^j。这里,我们常把离散信号原始输入数字序列的尺度作为 $1(2^{+0})$,其对应的分辨率为 $1(2^{-0})$,而其后的分析尺度则总是大于 $1(2^{+0})$为 $2^1,2^2,\cdots,2^n$,其对应的分辨率为 $2^{-1},2^{-2},\cdots,2^{-n}$。对连续函数 $f(x)\in L^2(R)$而言,其可能的尺度范围应当是 $0\to\infty,2^{-\infty}\to2^{+\infty}$。

2. 小波与多尺度分析

为了克服常规傅氏变换有时不能提取频域的局部特征，Gabor(1946)提出了窗口傅氏变换，即

$$F(\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}f(t)g(t-b)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t \tag{5.50}$$

并实现了时域局部化思想。

式中：$f(t)$是时间 t 的函数；$F(\omega)$是 $f(t)$的谱；$g(t-b)$是窗口函数。但是 $g(x)$一旦选定，它的大小、形状在各个不同的时(空)域位置处都是一样的，不能满足高频和低频信号对窗口大小的不同要求。为此，Goupillaud，Grossmann 和 Morlet(1984)引进了小波的概念。

根据定义，我们称满足条件

$$\int_{-\infty}^{\infty}|\hat{\psi}(w)|^2|w|^{-1}\mathrm{d}w<\infty \tag{5.51}$$

的函数 $\psi(t)\in L^2(R)$为小波函数或母小波，$\hat{\psi}$ 是 $\psi(t)$的傅氏变换。

若

$$\psi_{a,b}(t)=|a|^{-1/2}\psi\left(\frac{t-b}{a}\right)\qquad(a,b\in R,a\neq 0) \tag{5.52}$$

则称 $\psi_{a,b}$为由母小波 ψ 生成的依赖于参数 a,b 的连续小波；a 称为尺度变量；b 为位置变量。显然，连续小波是基于仿射群 $at+b$ 的，它们是通过母小波ψ的平移和伸缩而得到的。当尺度变量a的改变，是$\psi_{a,b}(t)$相对于母小波 $\psi(t)$而言，发生了伸缩改变；而当位置变量 b 发生变化时，$\psi_{ab}(t)$相对于母小波 $\psi(t)$而言发生了平移。与窗口傅氏变换相比，a 相当于 w，而这里的 b 与窗口函数 $g(t-b)$中的 b 相当。如果，将 $\psi_{a,b}(t)$视为窗函数，则通过 a 的压缩和 b 的平移，可以得到窗口大小和形状各异、但波的振荡数均相同的窗函数，b 实际上对应于物理空间中的实际位置；$|a|^{-1/2}$是一个归一化因子，它使所有的小波具有相同的模。因此，所有的连续小波 $\psi_{a,b}(t)$都具有相同的能量。

对于任一 $f(t)\in L^2(R)$的函数而言，定义

$$w_f(a,b)=\langle f,\psi_{a,b}\rangle=|a|^{-1/2}\int_{-\infty}^{\infty}f(t)\overline{\psi\left(\frac{t-b}{a}\right)}\mathrm{d}t \tag{5.53}$$

为其小波变换，式中：$\langle f,\psi_{a,b}\rangle$表示内积；而$\overline{\psi\left(\frac{t-b}{a}\right)}$表示$\psi\left(\frac{t-b}{a}\right)$的共轭。

其逆变换公式为：

$$f(t)=c_\psi^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}w_f(a,b)\psi_{a,b}(t)\frac{\mathrm{d}a\mathrm{d}b}{a^2} \tag{5.54}$$

式中：

$$c_\psi=\int_{-\infty}^{\infty}|\psi(w)|^2|w|^{-1}\mathrm{d}w \tag{5.55}$$

在实际计算中，常用其离散形式。若令：$a=2^{-j},b=2^{-j}\cdot i$ $(i,j\in Z)$，则有著名的二进小波，即

$$\psi_{ij}(t)=2^{j/2}\psi(2^jt-i)\qquad (i,j\in Z) \tag{5.56}$$

二进小波构成$L^2(R)$的一个正交基。利用ψ_{ji}可以将在无穷大处衰减得充分快的任意函数$f(t)$分解为：

$$f(t)=\sum_{j=-\infty}^{+\infty}\sum_{i=-\infty}^{+\infty}\langle f,\psi_{ji}\rangle\psi_{ji}(t),\qquad (j,i\in Z) \tag{5.57}$$

若设

$$D_{2j}=\sum_{i=-\infty}^{+\infty}\langle f,\psi_{ji}\rangle\psi_{ji}(t) \tag{5.58}$$

则(5.57)式可写为：

$$f(t)=\sum_{n=-\infty}^{+\infty}D_2n=\sum_{\substack{n=-\infty\\ \text{大尺度}}}^{-(j+1)}D_2n+\sum_{\substack{n=-j\\ \text{小尺度}}}^{+\infty}D_2n \tag{5.59}$$

上式中，等式右端第一项反映的是尺度大于2^j的信号分量之和，可以看作对应尺度2^j的$f(t)$之平滑部分；而等式右端第二项反映的是小于或等于2^j的信号分量之和，可以看作对应尺度2^j的$f(t)$的细节部分。

Mallat(1989)证明，对任意函数而言，上述多尺度逼近，存在唯

一的函数 $\psi(t)\in L^2(R)$称为尺度函数（或小波之父），且满足

$$f(t)=\sum_{i=-\infty}^{\infty}\langle f,\varphi_{ji}\rangle\varphi_{ji}(t)+\sum_{n=-j}^{\infty}\sum_{n=-\infty}^{\infty}\langle f,\psi_{ni}\rangle\psi_{ni}lt\rangle \tag{5.60}$$

上式中第一个和式是 2^j 尺度上 $f(t)$的一个光滑近似；而第二个和式是 2^j 尺度上 $f(t)$的细节部分之和。基于(5.60)式的分析方法则称为多尺度分析。

如果基于 2^j 尺度上对函数 $f(t)$的光滑近似属于空间 V_j，$f(t)$的细节部分属于 V_j 的正交补空间 W_j，则有：

$$V_{j-1}=V_j\oplus W_j=V_{j+1}\oplus W_{j+1}\oplus W_j=\cdots \qquad (j\in Z) \tag{5.61}$$

当 $j=1$ 时，

$$\begin{aligned}V_0&=V_1\oplus W_1=V_2\oplus W_2\oplus W_1\\&=V_3\oplus W_3\oplus W_2\oplus W_1=\cdots\end{aligned} \tag{5.62}$$

这种分解可以用图 5.8 表示。当我们把反问题从 V_j 空间分解为 $V_{j+1}\oplus W_{j+1}$时，由于 V_{j+1}空间仅包含 V_j 空间中的低波数成分，因此在 V_{j+1}空间中求得的解应是 V_j 空间中反问题在所有模型参数附近局部化的某种平均值解。又由于 V_{j+1}在 V_j 中的正交空间 W_{j+1}含有 V_j 空间中的高波数成分，因此在 V_{j+1}空间中不能得到分辨的小尺度特征，却能通过在 V_j 空间中的求解而得到分辨。这就是多尺度方法逐渐提高分辨率的原理。

3. *多尺度反演法*

多尺度反演算法可以表示为三个基本算子的操作过程：第一个算子将地球物理反问题从尺度零（小尺度）依次分解为尺度 1，2，…等大尺度的反问题；第二个算子求取各尺度上反问题的解；第三个算子将 $j(j\in Z)$尺度上的解嵌入尺度 $j-1$，并将其作为尺度 $j-1$ 上反问题寻优的起始点。为方便理解，下面以尺度 2 的情况加以说明。多尺度时，可以类推。

如设实测数据是按等间隔采样而获得的，设其采样间隔为 h，记其反问题所在空间为 E^h，其对应尺度为零。相应地 E^{2h}表示尺度为 1

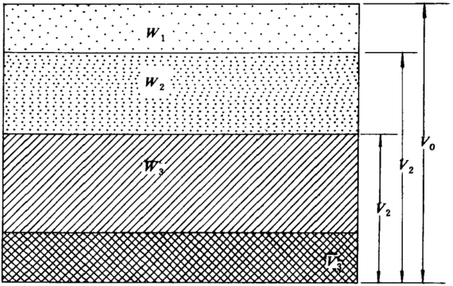


图 5.8　多尺度分解示意图

时反问题所在空间，其对应的数据相当于在 E^h 中以 $2h$ 为间隔，依某一准则进行重新采样而获得，这一过程可表示为：

$$I_h^{2h}: E^h \to E^{2h} \tag{5.63}$$

式中：I_h^{2h} 表示地球物理反问题从原空间 E^h（尺度为零）映射到 E^{2h}（尺度为 1）的多尺度分析算子。然后，在 E^{2h}空间求反问题的解，用算符 R^{2h}表示。第三个算子 I_{2h}^h将 E^{2h}空间经 R^{2h}得到的反问题的解嵌入 E^h 空间中，即：

$$I_{2h}^h: E^{2h} \to E^h \tag{5.64}$$

这种“嵌入”表示将 E^{2h}的解作为 E^h 中原反问题寻优的起始点。

设地球物理线性反演问题的数据方程为：

$$\underset{M\times N}{G}\ \underset{N\times 1}{m} = \underset{M\times 1}{e}$$

这里，有三种方法可实现多尺度分解反演。

第一种方法：

上式两端同乘以选择的小波（如 haar 小波或 Daubechies 小波）对应的变换矩阵 W，得：

$$WGm=We \tag{5.65}$$

式中采样点数 $M=2^j$，对应于尺度 2^{+j}　　$(j=1,2,\cdots,k)$。

将上式重新整理得：

$$\overline{G}m=\overline{e} \tag{5.66}$$

这里：$\overline{G}=WG$，　$\overline{e}=We$.

当 $j=1$ 时，$M=2$，反演两个数据的(5.65)方程，可以用前面讲述的各种方法求解，然后将其结果作为 $j=2$，即 4 个数据时反演问题的初始模型，再反演(5.65)，依次类推，计算 8 个数据，16 个数据…的反演问题，直至最小的尺度，即最大采样率时的反问题。这时的解，就是问题的最终解。

这种尺度分解算法的优点是只需对等式两端作一次小波变换，计算量小；缺点是尺度分解仅作用于数据 e 和核函数 G，对模型 m 无尺度运算概念。如 $M<N$，则分尺度的求解总是欠定问题，而欠定问题的求解又总要在一定假设条件下才能完成，这就造成不同尺度解之间继承性不能保证，不能充分利用多尺度分解的特点。

第二种方法：

$$We=WGW^{T}Wm=(W(WG)^{T})^{T}Wm \tag{5.67}$$

$$\overline{e}=\overline{G}\,\overline{m}$$

式中：$\overline{G}=(W(WG)^{T})^{T}$；　　$\overline{m}=Wm$

反演时与第一种方法一样，开始从大尺度开始，逐渐缩小尺度，直到最小尺度。

第三种方法：

$$e=GW^{T}Wm=(WG^{T})^{T}Wm=\overline{G}\overline{m} \tag{5.68}$$

式中：$\overline{G}=GW$

以上三种方法可分别适用于不同地球物理反演问题。

如前面所述，反演 2^j 尺度地球物理问题时，可以采用任何一种

合适的反演方法，如广义反演法等。

在 2^j 尺度上的解，用作下一尺度 2^{j-1} 的初始模型时，需要进行插值，即解的样点要加密一倍。从某种意义上讲，插值本身就是从 2^j 尺度对 2^{j-1} 尺度解的一种猜测，插值方法不同，结果也不一样。一般采用样点复制或线性插值的方法。

如前所述，非线性反演法，近十年来受到地球物理学家和应用数学家的广泛关注，发表了大量的论文和研究成果，本书中仅只简要地举它们中的一部分。由于非线性问题涉及广泛，不仅与大量数学、物理问题有关，而且也渗透着其他基础学科(如生物、化学、地质)和应用学科(如计算机科学、经济学…)的研究成果。可以说，非线性反演法的发展，是学科交叉、联合的反映。正由于它是一门全新的新兴学科，必然还有许多不足，理论不完备，方法不完善，效果还有待进一步提高。也就是说，还有许多问题需要开展深入的研究。

和线性反演法不同，一般说来，非线性反演法计算工作量都很大，要求并行机。因此，硬件不足，也往往成了非线性反演法前进的一大障碍。

虽然，比起线性反演法而言，非线性收敛到局部极小的可能性已大大减小，但实践证明，它并未从根本上解决迭代最终收敛于总体极小这一棘手问题。然而，它确实向既定目标——减小反演中的非唯一性范围，避免迭代中收敛于局部极小的情况发生——前进了一大步。

结 束 语

书已经写完了，但又觉得该说的话还没有言尽。如果把这几句该说的话放在正文，又似乎不妥，有文不对题之嫌，甚至影响已经确定的体系。所以，只好把几句言而未尽的话放在这里了。

在当今地球科学大发展的新时代，地球物理学对确定新的地球观提供了决定性的证据，起了十分重要的作用。而在寻找有用矿产（如石油、天然气以及各种金属和非金属矿），使之为人类服务方面，地球物理勘探又战功显赫。在这两方面反演理论都建立了不朽的功绩。然而，随着勘探的不断深入，待解决问题复杂性的加大，如埋深不断加大等，单一物探方法的反演工作已不完全适应新形势的要求，联合反演已迫在眉睫。因为我们借以寄居的这颗行星，本来就是千姿百态，千变万化的，不可能用一种地球物理特征去描述、或者准确地描述一种地质构造、一种矿产资源。为此，必须采取综合的战略。把单一地球物理资料的反演变为反映各种不同地球物理场或地球物理资料的联合反演，是当今地球物理资料反演的总趋势，也是发展的必然。我们提倡联合反演，但并不否认对单一资料反演方法深入研究的必要，特别是提高分辨力、减少非唯一性方面的研究的必要。

事实上，不少地球物理学家正在专心致志地从事这项研究工作。应该说，联合反演方法和单一资料的反演方法，原则上并无不同。不同的是，首先必须弄清哪些地球物理资料可以实现联合，达到反演的目的；其次，如何构制目标函数，才能达到最佳的联合效果；第三，选择何种限制条件，强加什么样的先验信息，才能最大限度地减小解的非唯一性，提高分辨力，以增强地球物理资料反演的地质效果。

当今地球物理资料反演的另一发展趋势是更多地注意非线性反

演方法的研究，这不仅是因为大多数地球物理问题都是非线性问题，对于非线性问题，最好的办法是采用非线性的方法去研究，而不是采用线性逼近或其他什么方法去研究。而且，非线性反演方法可以有效地减少解的非唯一性，尽量避免在反演迭代过程中，陷入目标函数的局部极小，因而，可以提高分辨力，增强反演的地质效果。

如前所述，非线性反演方法还处于研究阶段，尚未取得像线性反演法那样明显的地质成果。要在非线性反演研究方面取得成果，必须加强基础理论和应用两个方面的研究。根深才能叶茂，只有那些数理基础扎实，知识面广，地球物理反演功底雄厚的人，才可能在像非线性反演理论这样的交叉学科的研究方面取得成功。

此外，努力开展二维、三维地球物理资料的反演，是反演理论发展的另一个重要方向。事实上，一维地球物理模型是不存在的，它只是实际地球物理模型的一种十分粗略的近似。绝大多数地球物理模型，不是二维就是三维的。因此，开展这方面的研究就显得十分重要和完全必要。就地球物理资料反演现状而言，二维反演已开始应用于生产实践，三维反演仍处于研究和试验阶段，其原因不在于当今的反演方法和理论跟不上生产实践的发展，而在于硬件的发展跟不上形势，因为三维反演需要大量的计算时间和昂贵的计算费用。应该说，并行计算机的出现，为三维地球物理资料反演问题的普遍推广提供了必要的环境保证和硬件支撑。

为什么二、三维反演需要大量的计算时间呢?实验告诉我们，二、三维反演中真正耗费 CPU 的是正演计算。因为除极个别的简单二、三维地球物理模型的正演问题可以用解析公式计算外，一般都只能采用数值的方法：即有限差分、有限单元、积分方程和边界元法进行计算。即使计算一个简单的三维模型的正演问题，其计算时间也大得惊人。据报道，计算一个简单三维模型的 MT 正演问题，用 IBM-486 也得用去几十小时的时间。这不但增加了计算经费，更重要的是也使这种方法失去了实用价值，所以尽快使用到并行机是实现二、三维反

演的重要前提。

最后，还应该指出，到目前为止，对反演结果的评价，不管是线性评价还是非线性评价，在我国均未得到必要的重视，评价是反演理论的一个重要问题，应该得到应有的关注。

参 考 文 献

[1]Backus, G. E., 1970 a. Inference from inadequate and inaccurate dafa 1, proc. Nat. Acad. Sci. U. S. ,65,1~7.

[2]Backus, G. E., 1970 b. Inference from inadequate and inaccurate data 2, proc. Nat. Acad. Sci. U. S. ,65,281~287.

[3]Backus, G. E., 1970 c, Inference from inadequate and inaccurate data 3, proc. Nat. Acad. Sci. U. S. ,67,282~289.

[4]Backus, G. E., 1972. Inference from inadequate and inaccurate data, in mathematical problems in the geophysical sciences, American Mathematical Society, Providence, R. I.

[5]Backus, G. E. and Gilbert, F., 1967. Numerical applications of a formalism for geophysical inverse problems, Geophys. J. R. Astron. Soc., 13, 247~276.

[6]Backus, G. E. and Gilbert, F., 1968. The resolving power of gross earth data, Geophys. J. R. Astron. Soc., 16, 169~205.

[7]Backus, G. E. and Gilbert, F., 1970, Uniqueness in the inversion of inaccurate gross earth data, phil. Trans. R. Soc. ,A266, 123~192.

[8]Parker, R. L., 1970, The inverse problem of electrical conductivity in the mantle, Geophys. J. R. Astron. Soc. ,22,121~138.

[9]Parker, R. L. ,1977 a, Understanding inverse theory, Ann. Rev. Earth. Plan. Phys. ,5,35~64.

[10]Parker, R. L., 1977 b, The frechet derivative for the one dimentional electromagnetic induction problem, Geophys. J. R. Astron. Soc. ,49,543~547.

[11]Parker, R. L., 1977 c, Linear inference and underparameterized models, Rev. Geophys. Space phys. ,15,446~455.

[12]Parker,R. L. ,1971,The determination of seamount magnetism,Geophys. J. R. Astron. Soc. ,24,321～324.

[13]Parker,R. L. ,1972,Inverse theory with grossly inadequate data,Geophys. J. R. Astron. Soc. ,29,123～138.

[14]Parker, R. L. , 1980, The inverse problem of electromagnetic induction. existence and construction of solutions based on incomplete data: J. Geophys. Res. ,85,4421～4428.

[15] Parker, R. L. , 1982, The existence of a region inaccessible to magnetotelluric sounding,Geophys. J. R. Astron. Soc. ,68,165～170.

[16]Parker,R. L. ,1983,The magnetotelluric inverse problem,Geophys. Surv. , 6,5～25.

[17]Parker,R. L. ,and McNutt,M. K. ,1980,Statistics for the one-norm misfit measure. J. Geophys. Res. ,85,4429～4430.

[18] Parker, R. L. , and Whaler, K. A. , 1981, Numerical methods for establishing solutions to the inverse problem of electromagnetic induction. J. Geophys. Res. ,86,9574～9584.

[19] Oldenburg, D. W. , 1979, One-dimensional inversion of natural source magnetotelluric observations, Geophysics,44,1218～1244.

[20] Oldenburg, D. W. , 1983, Funnel functions in linear and non-linear appraisal,J. Geophys. Res. ,88,7387～7398.

[21]Oldenburg,D. W. ,1984,An introduction to linear inverse theory,IEEE, GE-22,664～674.

[22]Oldenburg,D. W. ,Whittall,K. P. ,and Parker,R. L. ,1984,Inversion of ocean bottom magnetotelluric data revisited. J. Geophys. Res. 89,1829～1833.

[23] Oldenburg, D. W. , 1981, A comprehensive solution to the linear deconvolution problem,Geophys. J. R. Astron. Soc. ,65,331～357.

[24]Oldenburg,D. W,1982,Multichannel appraisal deconvolution. Geophys. J. R. Astron. Soc. ,69,405～414.

[25]Oldenburg,D. W. ,1976,Calculation of Fourier transforms by the Backus-Gilbert method,Geophys. J. R. Astron . Soc. ,44,413～431.

[26] Oldenburg, D. W., 1981, Conductivity structure of the oceanic upper mantle beneath the Pacific plate, Geophys. J. R. Astron. Soc., 65, 359～394.

[27] Oldenburg, D. W., Levy, S., and Whittall, K. P., 1981, Wavelet estimation and deconvolution, Geophysics, 46, 1528～1542.

[28] Jupp, D. L., and Vozoff, K., 1975, Stable iterative methods for the inversion of geophysical data, Geophys. J. R. Astron., Soc., 42, 957～976.

[29] Kunetz, G., 1972, Processing and interpretation of magnetotelluric sounding, Geophysics, 37, 1005～1021.

[30] Lanczos, C., 1961, Linear differential operators, D. Van Nostrand Co.

[31] Menke, M., 1984, Geophysical data analysis: Discrete inverse theory, Academic Press, Inc.

[32] Smith, J. T., and Booker, J. R., 1988, Magnetotelluric inversion for minimum structure, Geophysics, 53, 1565～1576.

[33] Tarantola, A., 1987, Inverse problem theory: methods for data fitting and model parameter estimation, Elsevier Science Publishing Company Inc.

[34] Rothman, D. H., 1985, Non-linear inversion, statistical mechanics, and residual statics estimation, Geophysics, 50, 2784～2796.

[35] Rothman, D. H., 1986, Automatic estimation of large residual statics correction, Geophysics, 51, 332～346.

[36] Sen, M. K., Bhattacharya, B. B., and Stoffa, P. L., 1993, Non-linear inversion of resistivity sounding data, Geophysics, 58, 496～507.

[37] Sen, M. K., Stoffa, P. L., 1991, Non-linear one-dimensional seismic waveform inversion using simulated annealing. Geophysics, 56, 1624～1638.

[38] Sen, M. K., Stoffa, P. L., 1992, Rapid sampling of model space using genetic algorithms: Examples from seismic waveform inversion, Geophys. J. Int., 108, 281～292.

[39] Stoffa, P. L., Sen, M. K., 1991, Non-linear multiparameter optimization using genetic algorithms: Inversion of plane-wave seismograms, Geophysics, 56, 1794～1810.

[40] Press, F. , 1968, Earth models obtained by Monte-Carlo inversion, J. Geophys. Res. ,73,5223～5234.

[41] Weidelt, P. , 1972, The inverse problem of geomagnetic induction, Zeit. Geophys. ,38,257～289.

[42] Levy, S. , Oldenburg, D. , and Wang, J. , 1988, Subsurface imaging using magnetotelluric data, Geophysics, 53, 104～117.

[43] Wiggins, R. A. , 1972, The general inverse problem: implication of surface phys. ,10,251～285.

[44] Whittall, K. P. , and Oldenburg, D. W. , 1992, Inversion of magnetotelluric data for a one-dimensional conductivity, Society of Exploration Geophysicists.

[45] Zhang, Y. , and Paulson, K. V. , 1997, Magnetotelluric inversion using regularized Hopfield neural networks, Geophysical prospecting, 725～743.

[46] Goupillaud, A. , Grossmann, A. , and Morlet, J. , 1984, Cycleoctave and related transforms in seismic signal analysis, Geoexploration, 23, 85～102.

[47] Hopfield, J. J. , and Tank, D. W. , 1985, Neural computation of decisions in optimization problems, Biological cybernetics, 52, 141～152.

[48] Gabor, D. , 1946, Theory of communication, J. IEE, 93, 429～457.

[49] Mallat, S. , 1989, A theory of multiresolution signal decomposition: the wavelet representation ,IEEE Trans. Patte Anal. Machine Intell. ,11,674～693.

[50]王家映,1985,大地电磁拟地震解释法,石油地球物理勘探,20,66～79。

[51]王家映,1993,石油电法勘探。北京:石油工业出版社。

[52]王家映,1988,地球物理学。武汉:中国地质大学出版社。

[53]王家映,1995,大地电磁拟地震解释法。北京:石油工业出版社。

[54]欧登伯格,王家映,1984,地球物理学中的反演理论,地球科学,第9卷,第3期,133～147。

[55]杨文采,1989,地球物理反演和地震层析成像。北京:地质出版社。

[56]许 云,1981,地震层状构造理论。北京:石油工业出版社。

[57]姚 姚,1995,地球物理非线性反演模拟退火法的改进,地球物理学报,38,643～650。

[58]靳　蕃,范俊波,谭永东,1991,神经网络与神经计算机:原理.应用。成都:西南交通大学出版社。
[59]崔锦泰(美)著,程正兴译,1995,小波分析导论。西安:西安交通大学出版社。
[60]李世雄,刘家琦,1994,小波变换和反演数学基础。北京:地质出版社。
[61]李庆忠,1994,走向精确勘探的道路。北京:石油工业出版社。
[62]杨文采,1993,地震道的非线性混沌反演——理论和数值试验,地球物理学报,36。
[63]师学明,王家映,1988,一维层状介质大地电磁模拟退火反演法,地球科学V.23,NO.5,P542~545。